

THE
QUARTERLY
JOURNAL OF
ECONOMICS

FOUNDED 1886

Editor

ARTHUR SMITHIES

VOLUME LXXII

PUBLISHED BY HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS

1958

PRINTED IN UNITED STATES OF AMERICA

THE MURRAY PRINTING COMPANY
FORGE VILLAGE, MASSACHUSETTS

CONTENTS FOR VOLUME LXXII

WRITERS	PAGE
ALHADEFF, CHARLOTTE P. AND ALHADEFF, DAVID A. The Struggle for Commercial Bank Savings	1
ALHADEFF, DAVID A. AND ALHADEFF, CHARLOTTE P. The Struggle for Commercial Bank Savings	1
BANNER, PAUL H. Competition, Credit Policies, and the Captive Finance Company	241
BATOR, FRANCIS M. The Anatomy of Market Failure	351
BLACK, JOHN D. The Extremities of Current Agricultural Policy Proposals	307
BLAUG, MARK. The Classical Economists and the Factory Acts — A Re-Examination	211
BUTTRICK, JOHN. A Note on Professor Solow's Growth Model	633
CAMPBELL, COLIN D. Investments in United States Government Securities by Nonfinancial Corporations, 1952-1956: Comment	292
CULBERTSON, J. M. Culbertson on Interest Structure: Reply	607
DICKINSON, Z. CLARK. A Letter from Adam Smith	157
ELLIOTT, JOHN E. Economic Planning Reconsidered	55
ELLIS, HOWARD S. Accelerated Investment as a Force in Economic Development	485
ENKE, STEPHEN. Controlling Consumers in the Event of Future Wars	558
GABOR, ANDRÉ AND PEARCE, I. F. The Place of Money Capital on the Theory of Production	537
GOLAY, FRANK H. Commercialism and Economic Nationalism	574
HAGEN, EVERETT E. An Economic Justification of Protectionism	496
HINSHAW, RANDALL. The Effect of Devaluation on the Price Level: Further Comment	616
HIRSCHMAN, ALBERT O. AND SIRKIN, GERALD. Investment Criteria and Capital Intensity Once Again	469
HIRSHLEIFER, JACK. Peak Loads and Efficient Pricing: Comment	451
HOUTHAKKER, H. S. Peak Loads and Efficient Pricing: Further Comment	463
ISAAC, J. E. The Function of Wage Policy: The Australian Experience	115
JOHNSON, HARRY G. Tariffs and the Balance of Payments: Comment	301
LECAILLON, JACQUES AND MARCHAL, JEAN. Is the Income of the "Cadres" a Special Class of Wages?	166
LEVIN, HARVEY J. Economic Structure and the Regulation of Television	424
MANSFIELD, EDWIN AND WEIN, HAROLD H. A Study of Decision-Making Within the Firm	515
MARCHAL, JEAN AND LECAILLON, JACQUES. Is the Income of the "Cadres" a Special Class of Wages?	166
MINSKY, HYMAN P. Investments in United States Government Securities by Nonfinancial Corporations, 1952-1956: Reply	297
MODIGLIANI, F. AND WEINGARTNER, H. M. Forecasting Uses of Anticipatory Data on Investment and Sales	23
MOSES, LEON N. Location and the Theory of Production	259
NERLOVE, MARC. Adaptive Expectations and Cobweb Phenomena	227

WRITERS	PAGE
NICHOLSON, HOWARD W. Motor Carrier Costs and Minimum Rate Regulation	139
OZGA, S. A. Tariffs and the Balance of Payments: Reply	302
PEARCE, I. F. AND GABOR, ANDRÉ. The Place of Money Capital in the Theory of Production	537
PURSELL, G. Unity in the Thought of Alfred Marshall	588
SEGRÉ, CLAUDIO. Monetary Surveys and Monetary Analysis	98
SHEAHAN, JOHN. International Specialization and the Concept of Balanced Growth	183
SIRKIN, GERALD AND HIRSCHMAN, ALBERT O. Investment Criteria and Capital Intensity Once Again	469
SMITHIES, ARTHUR. The Rockefeller Report: Implications for Fiscal and Monetary Policy	472
SOHMEN, EGON. The Effect of Devaluation on the Price Level	273
SOHMEN, EGON. The Effect of Devaluation on the Price Level: Reply	627
SOSNICK, STEPHEN H. A Critique of Concepts of Workable Competition	380
STEINER, PETER O. Peak Loads and Efficient Pricing: Reply	465
STEVENS, CARL M. On the Theory of Negotiation	77
TOBIN, JAMES. The Business Cycle in the Post-War World: A Review	284
VALAVANIS, STEFAN. Traffic Safety from an Economist's Point of View	477
VANEK, JAROSLAV. The Effect of Devaluation on the Price Level: Comment	614
WEHRLE, L. S. Culbertson on Interest Structure: Comment	601
WEIN, HAROLD H. AND MANSFIELD, EDWIN. A Study of Decision-Making Within the Firm	515
WEINGARTNER, H. M. AND MODIGLIANI, F. Forecasting Uses of Anticipatory Data on Investment and Sales	23
WRIGHT, A. LL. The Rate of Interest in a Dynamic Model	327
WRIGHT, DAVID McCORD. What <i>Is</i> The Economic System?	198

SUBJECTS

Accelerated Investment as a Force in Economic Development. By Howard S. Ellis	485
Adaptive Expectations and Cobweb Phenomena. By Marc Nerlove	227
An Economic Justification of Protectionism. By Everett E. Hagen	496
The Anatomy of Market Failure. By Francis M. Bator	351
The Business Cycle in the Post-War World: A Review. By James Tobin	284
The Classical Economists and the Factory Acts — A Re-Examination. By Mark Blaug	211
Commercial Policy and Economic Nationalism. By Frank H. Golay	574
Competition, Credit Policies, and the Captive Finance Company. By Paul H. Banner	241
Controlling Consumers in the Event of Future Wars. By Stephen Enke	558
A Critique of Concepts of Workable Competition. By Stephen H. Sosnick	380
Culbertson on Interest Structure: Comment. By L. S. Wehrle	601
Culbertson on Interest Structure: Reply. By J. M. Culbertson	607
Economic Planning Reconsidered. By John E. Elliott	55

SUBJECTS	PAGE
Economic Structure and the Regulation of Television. By Harvey J. Levin	424
The Effect of Devaluation on the Price Level. By Egon Sohmen	273
The Effect of Devaluation on the Price Level: Comment. By Jaroslav Vanek	614
The Effect of Devaluation on the Price Level: Further Comment. By Randall Hinshaw	616
The Effect of Devaluation on the Price Level: Reply. By Egon Sohmen .	627
The Extremities of Current Agricultural Policy Proposals. By John D. Black	307
Forecasting Uses of Anticipatory Data on Investment and Sales. By F. Modi- gliani and H. M. Weingartner	23
The Function of Wage Policy: The Australian Experience. By J. E. Isaac .	115
International Specialization and the Concept of Balanced Growth. By John Sheahan	183
Investment Criteria and Capital Intensity Once Again. By Albert O. Hirschman and Gerald Sirkin	469
Investments in United States Government Securities by Nonfinancial Cor- porations, 1952-1956: Comment. By Colin D. Campbell	292
Investments in United States Government Securities by Nonfinancial Cor- porations, 1952-1956: Reply. By Hyman P. Minsky	297
Is the Income of the "Cadres" A Special Class of Wages? By Jean Marchal and Jacques Lecaillon	166
A Letter from Adam Smith. By Z. Clark Dickinson	157
Location and the Theory of Production. By Leon N. Moses	259
Monetary Surveys and Monetary Analysis. By Claudio Segré	98
Motor Carrier Costs and Minimum Rate Regulation. By Howard W. Nicholson	139
A Note on Professor Solow's Growth Model. By John Buttrick	633
On the Theory of Negotiation. By Carl M. Stevens	77
Peak Loads and Efficient Pricing: Comment. By Jack Hirshleifer	451
Peak Loads and Efficient Pricing: Further Comment. By H. S. Houthakker	463
Peak Loads and Efficient Pricing: Reply. By Peter O. Steiner	465
The Place of Money Capital in the Theory of Production. By André Gabor and I. F. Pearce	537
The Rate of Interest in a Dynamic Model. By A. Ll. Wright	327
The Rockefeller Report: Implications for Fiscal and Monetary Policy. By Arthur Smithies	472
The Struggle for Commercial Bank Savings. By David A. Alhadeff and Charlotte P. Alhadeff	1
A Study of Decision-Making Within the Firm. By Edwin Mansfield and Harold H. Wein	515
Tariffs and the Balance of Payments: Comment. By Harry G. Johnson .	301
Tariffs and the Balance of Payments: Reply. By S. A. Ozga	302
Traffic Safety from an Economist's Point of View. By Stefan Valavanis .	477
Unity in the Thought of Alfred Marshall. By G. Pursell	588
What Is The Economic System? By David McCord Wright	198

THE
QUARTERLY
JOURNAL OF
ECONOMICS

FOUNDED 1886

Vol. LXXII

February, 1958

No. 1

THE STRUGGLE FOR COMMERCIAL BANK SAVINGS

DAVID A. ALHADEFF and CHARLOTTE P. ALHADEFF 1

FORECASTING USES OF ANTICIPATORY DATA ON INVESTMENT

AND SALES.....F. MODIGLIANI and H. M. WEINGARTNER 23

ECONOMIC PLANNING RECONSIDERED.....JOHN E. ELLIOTT 55

ON THE THEORY OF NEGOTIATION.....CARL M. STEVENS 77

MONETARY SURVEYS AND MONETARY ANALYSIS..CLAUDIO SEGRÉ 98

THE FUNCTION OF WAGE POLICY: THE AUSTRALIAN

EXPERIENCE.....J. E. ISAAC 115

MOTOR CARRIER COSTS AND MINIMUM RATE REGULATION

HOWARD W. NICHOLSON 139

Recent Publications..... 153

PUBLISHED BY HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS

The Quarterly Journal of Economics

Published by Harvard University

Editor

EDWARD H. CHAMBERLIN

Associate Editors

ABRAM BERGSON
JOHN D. BLACK*
J. K. BUTTERS
T. N. CARVER*
ROBERT DORFMAN
JAMES S. DUESENBERY
JOHN T. DUNLOP
JOHN KENNETH GALBRAITH
ALEXANDER GERSCHENKRON
GOTTFRIED HABERLER
ALVIN H. HANSEN*

* Emeritus

SEYMOUR E. HARRIS
CARL KAYSEN
WASSILY LEONTIEF
EDWARD S. MASON
GUY H. ORCUTT
SUMNER H. SLICHTER
ARTHUR SMITHIES
OVERTON H. TAYLOR
ABBOT P. USHER*
JOHN H. WILLIAMS*
EDWIN B. WILSON*

Assistant Editor

CATHERINE RUGGLES GERRISH

Books, periodicals, and manuscripts to be addressed, THE QUARTERLY JOURNAL OF ECONOMICS, M-12, Littauer Center, Cambridge 38, Mass.

Business letters to be addressed, HARVARD UNIVERSITY PRESS, 79 Garden Street, Cambridge 38, Mass.

Subscriptions \$5.00 a year. Three years, \$14.00. Single copies (except those published in two parts) \$1.50, postpaid. A special subscription rate of \$3.00 is offered to students. Such special subscriptions may be for one, two, three, or four years. Those desiring to take advantage of this reduced rate should include with their remittance a statement from an officer of their college or university that they are *at that time* "regularly enrolled students who have not yet received the Doctor's degree."

CONTENTS FOR NOVEMBER, 1957

THE TERM STRUCTURE OF INTEREST RATES.....	John M. Culbertson
LABOR FORCE AND INDUSTRIAL DEVELOPMENT IN SOVIET GERMANY..	Wolfgang F. Stolper
THE MEANING OF THE FITTED COBB-DOUGLAS FUNCTION.....	E. H. Phelps Brown
SOME NOTES ON THE CHOICE OF CAPITAL-INTENSITY IN DEVELOPMENT PLANNING	Amartya Kumar Sen
PEAK LOADS AND EFFICIENT PRICING.....	Peter O. Steiner
KARL MARX AND SAY'S LAW.....	Bernice Shoul
TARIFFS AND THE BALANCE OF PAYMENTS.....	S. A. Ozga
Recent Publications	

COPYRIGHT, 1958, BY THE PRESIDENT AND FELLOWS OF HARVARD COLLEGE
Entered as Second-Class Mail Matter. Acceptance for mailing at special postage rate provided for by
Section 1103, Act of October 3, 1917, authorized on July 31, 1918
Published by the Harvard University Press, Cambridge, Mass., U.S.A.
London: Oxford University Press

THE QUARTERLY JOURNAL OF ECONOMICS

Vol. LXXII

February, 1958

No. 1

THE STRUGGLE FOR COMMERCIAL BANK SAVINGS

By DAVID A. ALHADEFF AND CHARLOTTE P. ALHADEFF

I. Introduction, 1. — II. The relative decline of commercial banks, 2. — III. Sources of comparative competitive strength, 11. — IV. Conclusions, 21.

I. INTRODUCTION

Our changing financial structure. The increasing importance of nonbank financial intermediaries (like life insurance companies, savings and loan associations, pension funds, etc.) has raised serious questions for monetary control, for the savings-investment process, for the allocation of resources, for the operations of the money markets, and, indeed, for the validity of monetary theory.¹ It is for these reasons that "a basic re-examination of our entire monetary and financial networks" is being urged in many quarters.² Effective public policy must be based on an understanding of the forces which are transforming our financial structure. One of the most important changes is the relative decline of the commercial banks in the savings field. The main argument of this paper is directed to the decline of commercial bank savings compared with that very rapidly growing financial intermediary, savings and loan associations (SLA).³ It is the purpose of this paper to discover the reasons for this development and to assess the comparative competitive strength of commercial banks.

The major legislation affecting commercial banks — especially the Federal Reserve Act — was passed during a period when com-

1. Cf., for example, the papers grouped under the title, "Institutional Aspects of Savings and Investment," *American Economic Review*, XLIV (May 1954), 440-85. Cf. also, Warren L. Smith, "On the Effectiveness of Monetary Policy," *American Economic Review*, XLVI (Sept. 1956), 588-606.

2. Cf., for example, the testimony of Elliot V. Bell, *Hearings before the Subcommittee on Economic Stabilization of the Joint Economic Committee*, 84th Congress, 2d Session, p. 4.

3. For a description of the salient characteristics of SLA see Harold W. Torgerson, "Developments in Savings and Loan Associations, 1945-1953," *Journal of Finance*, IX (Sept. 1954), 283-97.

mercial banks dominated the financial scene. New financial legislation is being debated to resolve issues like branch offices for commercial banks and for SLA, wider investment powers for SLA, revision of tax treatment for commercial banks and SLA, inclusion of SLA under some kind of reserve requirement together with commercial banks, additional power to control SLA lending, etc. These measures are concerned not only with monetary control but also with equalizing the competitive position of commercial banks and SLA. The future development of our financial structure can be profoundly affected by the legal framework. We hope that our analysis of the present relationship between banks and SLA will contribute to an understanding of the probable effects of the proposed measures.

Our analysis is limited to commercial banks and SLA, because these are now the most important institutional competitors for the public's savings in the general form of savings deposits.⁴ In addition, both institutions devote a substantial portion of the savings dollar to mortgage investment. In other words, they compete both in the factor market (savings deposits) and in the product market (mortgage loans).

II. THE RELATIVE DECLINE OF COMMERCIAL BANKS

Rapid growth of SLA. The shifting position of commercial banks and SLA in attracting savings deposits⁵ can be quickly summarized. The commercial bank share of individuals' savings accounts (including mutual savings bank deposits) was higher at the end of 1945 than in December 1939; the SLA share was unchanged.⁶ Between December 31, 1945, and December 31, 1955, individuals' savings accounts increased \$54 billion. SLA attracted \$25 billion; commercial banks, only \$16 billion. During this same period, the SLA share of individuals' savings accounts increased from 14 to 30 per cent, while the commercial bank share fell from 57 to 43 per cent. During 1955 alone individuals' savings accounts increased slightly more than

4. Mutual savings banks are ignored for two reasons. First, they are highly concentrated regionally (especially in New England), and operate in only seventeen states. Hence they do not have the same national interest as SLA, which are established in every state. Second, during the past ten years mutual savings banks have grown less rapidly than SLA. Accordingly they do not present a comparable threat to the long-established position of commercial banks in the savings field.

5. "Savings accounts" (or "savings deposits") refers both to time and savings deposits in commercial banks and to share accounts in SLA.

6. *Savings and Loan Fact Book, 1955*, p. 10. During this same period, the commercial bank share increased from 51 to 57 per cent.

\$8 billion; 61 per cent was placed in SLA, but only 15 per cent in commercial banks.⁷

It should be mentioned, too, that SLA growth has been a national phenomenon. Between December 31, 1949 and December 31, 1954, the ratio of the volume of SLA to commercial bank savings accounts

TABLE I

EFFECTIVE RATES OF RETURN ON TIME DEPOSITS AND SHARE ACCOUNTS,
1948-1955

Year	Member SLA ¹ (%)	Insured Commercial Banks (%)
1948	2.44	.90
1949	2.52	.91
1950	2.52	.94
1951	2.59	1.03
1952	2.70	1.15
1953	2.81	1.24
1954	2.88	1.32
1955	2.94	1.38

Sources: SLA rates were computed from data in Federal Home Loan Bank Board, *Combined Financial Statements*, 1948-1955. Dividends were divided by average (December 31 of each year and the preceding year) share capital. A rate series for all SLA is virtually identical to the member SLA series in the table. The former can be computed from data in *Savings and Loan Fact Book*, 1956, pp. 12 and 52.

Commercial bank rates are from FDIC, *Annual Report*, 1955, p. 136. The deposit figures for the FDIC are an average of figures reported at the beginning, middle, and end of each year.

1. In some states SLA are permitted to issue permanent stock. The amount of permanent stock for member SLA is insignificant compared with share accounts. To include permanent stock in the rate computation would generally lower the effective dividend by only .01.

increased more than 100 per cent in ten states, between 50 and 100 per cent in twenty-six states, between zero and 50 per cent in eleven states, and declined in only one state.⁸

High dividend rates and aggressive salesmanship. Different interest rates are the most popular explanation for different growth rates of savings in SLA and commercial banks. Table I shows that in every year between 1948 and 1955 effective interest⁹ on share accounts was substantially higher than on time deposits. The importance of rate

7. Computed from figures in Federal Home Loan Bank Board, *Investments of Individuals in Savings Accounts, U. S. Savings Bonds, and Life Insurance Reserves*, 1955.

8. The information for all states was computed from data in U. S. Savings and Loan League, *Savings and Loan Annals*, 1950, pp. 92-93, FDIC, *Operating Insured Commercial and Mutual Savings Banks — Assets and Liabilities, December 31, 1949*, pp. 12-60, and *Savings and Loan Fact Book*, 1956, p. 13. SLA figures include all SLA; commercial bank figures are for insured commercial banks only. The bank data for 1954 are based on June 1955 figures. Hence, the calculations in the text are only approximate, but probably accurate within the broad ranges cited.

9. Comprehensive information about nominal interest rates on savings accounts is not available.

of return on the placement decisions of savers has been stressed by economists,¹ bankers,² savers,³ and SLA managers.⁴ Indeed, both the professional and lay explanation for the postwar growth of SLA is a combination of the interest rate difference and aggressive salesmanship by SLA. Certainly these are the most obvious factors; certainly, too, the interest rate difference was essential for SLA to maintain their market share between 1939 and 1945. But as an explanation for the phenomenal *relative* decline of commercial bank savings in the postwar years, the popular explanation is superficial; it is half wrong, and it is seriously incomplete in the remainder. It is the purpose of this section to reveal the error and inadequacy of the popular explanation and to propose our own explanation for the relative decline of commercial banks in the savings field.

The popular explanation is superficial as economic analysis, because it leaves many important questions unanswered. For example, is the commercial bank decline owing to an inability or an unwillingness to compete more effectively for the public's savings? Are commercial banks high-cost or low-cost producers, and why? Do commercial banks earn more or less than SLA on savings deposits, and why? In short, what is the basic competitive relationship between banks and SLA?

Low interest rates — necessary but not sufficient. The popular explanation is half wrong, because the interest difference between time deposits and share accounts cannot explain the dramatic shift of market shares during the postwar period. During the period 1939–1945, as we have shown, the commercial bank share of total savings deposits increased; the SLA share was unchanged. This pattern was sharply reversed during the postwar period, and relative interest rates actually moved inversely with market shares. Table II compares relative growth rates with effective relative interest rates between 1948 and 1955. It shows that the ratio of member share accounts to time deposits increased from 29 per cent to almost 67 per cent between 1948 and 1955. By contrast, the rate paid by SLA was 171 per cent higher than the commercial bank rate in 1948, but only 113 per cent higher in 1955. In other words, *the rapid relative decline of commercial bank savings took place in spite of the relative*

1. For example, cf. Torgerson, *op. cit.*, p. 286.

2. Cf. Gaylord Freeman, *Savings and Loan Competition*, p. 16.

3. "When all savers were asked which type of institution [commercial and savings bank or SLA] they would prefer if the rate of return were the same, 90% chose a bank." From a survey of savers by the U. S. Savings and Loan League, reported in *Savings and Loan News*, April 1954, p. 19.

4. SLA rates are typically higher than commercial bank rates.

increase in interest rates on time deposits. For those who believe that savers are sometimes motivated more by absolute than by percentage rate differences, we have also included the absolute differences. The table reveals that absolute rate differences were virtually constant during the period studied. Nevertheless, the commercial bank

TABLE II

COMPARATIVE GROWTH AND INTEREST RATES, COMMERCIAL BANKS AND SLA, 1948-1955

Year	Share Capital Time Deposits (%)	SLA Dividend Rate Time Deposit Rate (%)	Absolute Difference Between Rates (%)
1948	29.0	271	1.54
1949	33.0	277	1.61
1950	37.2	268	1.58
1951	41.8	251	1.56
1952	46.7	235	1.55
1953	52.2	227	1.57
1954	58.8	218	1.56
1955	67.3	213	1.56

Sources and notes: The ratios of share capital to time deposits were computed from data in Federal Home Loan Bank Board, *Combined Financial Statements*, for each year, and *Federal Reserve Bulletin*, for each year. In both cases, the data are based on year-end figures, and include only time deposits of individuals, partnerships, and corporations.

Comparative interest rates and absolute differences are based on Table I.

Information concerning interest rates on time deposits is available only for insured commercial banks, and official information concerning interest-dividend rates is available only for member SLA. Accordingly, volume figures for time deposits are limited to insured commercial banks; for share accounts, to member SLA. If all SLA and all commercial banks had been included, the ratios of share capital to time deposits would have varied from 31.4 per cent in 1948 to 70.2 per cent in 1955.

share of savings declined rapidly. We are forced to conclude, despite the chorus of voices to the contrary, that the interest rate difference is discredited as the explanation for the relative decline of commercial bank savings.

Changing preference patterns. The popular explanation is seriously incomplete in the remainder — its stress on salesmanship — because aggressive salesmanship is only one of many factors which have brought about the shift in market shares. To develop this part of the analysis, we must examine the nature of savers' preferences for time deposits and share accounts, the variables which affect the preference patterns, the shapes of the functions, and the changes which have occurred in the foregoing.

The savings "package." The main components of the savings account "package" are safety, availability (liquidity), convenience, and the interest rate. The placement decision of a saver is made after a subjective evaluation of the component elements of the savings account packages and a determination of the relative importance of the components.

For the moment, we defer our discussion of any changes in the subjective evaluation by savers. We have already shown that interest rates on time deposits rose relatively to rates on share accounts between 1948 and 1955. As for the other components of the package, comparative safety and availability in both institutions probably has not changed materially during the period studied. Moreover, the minor changes which have occurred are probably not widely known among savers and can be ignored in this context. In some ways, e.g., services, SLA convenience to savers has been enhanced. In probably the most important aspect of convenience, viz., location, the SLA competitive position has not materially changed, at least so far as this change can be measured by number of offices.⁵ Except for interest rates, therefore, there has been no material, objective relative change in the components of the savings package offered by commercial banks and SLA. Moreover, in the subjective evaluation of most savers, commercial banks offer greater combined safety, availability, and convenience than SLA.⁶ One bank spokesman put it more vividly: "... our competitors [SLA] hold only one trump card — rate of dividend — while we hold the rest of the entire deck."⁷

As for the determination by savers of the relative importance of the components, a survey of commercial bank and SLA savers revealed that "... safety and availability of funds and convenience of location are more important to most savers than a higher rate of return."⁸ Since most savers believe commercial banks have a superior combination of the preferred components, it is not surprising that (commercial and savings) bank savers are almost four times as numerous as SLA savers.⁹

5. Between 1948 and 1954, the number of commercial banks and of SLA was fairly stable, although the ratio of banks to SLA declined slightly. (Cf. *Federal Reserve Bulletin*, Feb. 1955, p. 208, and Feb. 1951, p. 228. Cf. also *Savings and Loan Fact Book*, 1955, p. 39.) The number of branches for both was far less stable, however. Commercial bank branches increased from 4,349 to 6,309, whereas federal SLA branches rose from 55 to 262. (Cf. *Federal Reserve Bulletin*, Feb. 1955, p. 208, and Feb. 1951, p. 228. Cf. also *Hearings before a Subcommittee of the Committee on Banking and Currency*, 84th Congress, 1st Session, on S. 972, pp. 6, 45, and 69.) Information about state SLA branches is not available. However, it seems unlikely that the rate of branch formation was higher among state SLA than among federal SLA, because state laws about branches are generally more restrictive than the federal law. Moreover, state SLA are mostly smaller than federal SLA. Even if we assume that the number of state SLA branches changed in the same proportion as federal SLA branches, the ratio of all commercial bank offices to all SLA offices would not have altered significantly during this period.

6. Cf. survey reported in *Savings and Loan News*, Aug. 1956, p. 25.

7. Cf. *Banking*, Mar. 1953, p. 96.

8. *Savings and Loan Annals*, 1953, p. 51.

9. From the survey in *Savings and Loan Annals*, 1953, p. 46. A more recent

The preference function. The saver's placement decision between share accounts and time deposits can be expressed in indifference terminology. For each saver, a set of indifference curves shows the substitutability of share accounts and time deposits. Another set of curves on the preference map shows what any level of savings would be worth after one year, given the relative rates of interest. The distribution of the individual's savings account is determined by the highest indifference curve that can be attained for any given volume of savings.

The marginal rate of substitution between share accounts and time deposit accounts reflects primarily subjective risk estimates of savers. Since most savers believe banks are both safer and more liquid than SLA, the individual saver would require more than proportional compensation of share accounts for a given reduction in (preferred) time deposit accounts, i.e., the indifference curves are negatively inclined and convex to the origin.¹ For our purposes, a determination of the exact shape of the indifference curves, beyond the fact of their negative slope, is not critical. Since most small savers (say, \$1,000 and under)² do not split their accounts,³ it is a sufficiently good approximation for our purposes to posit a constant marginal rate of risk substitution along the indifference curves for small savers.⁴ In fact, since splitting of accounts is not widely practiced even among large savers, the possibility will be ignored in this paper.⁵ Thus, when the marginal rate of risk substitution is smaller than the slope of the rate of return curve, savers will save exclusively in SLA, whereas, under the reverse relationship, they would save exclusively in commercial banks.

Even though we posit linear indifference curves, it should be

survey by the same organization showed a slightly lower ratio of bank savers to SLA savers. Cf. *Savings and Loan News*, May 1956, p. 37.

1. Moreover, the slopes of the indifference curves would not generally equal or be less than -1 , with time deposits on the ordinate.

2. Small savers are in a majority at both SLA and banks. Cf. *Savings and Loan News*, April 1954, p. 19. In 1953, 81 per cent of (commercial and savings) bank savers and 73 per cent of SLA savers had savings accounts of less than \$2,000. Cf. *Savings and Loan Annals*, 1953, p. 47.

3. In a 1955 survey of savers, only 6 per cent of savers with family incomes of less than \$4,000 a year split their savings between banks and SLA. Cf. *Savings and Loan News*, May 1956, p. 37.

4. Alternatively, the absence of widescale splitting may be owing to discontinuities at either end of the indifference curves, because of the nuisance of having two accounts, one necessarily quite small. Decisions falling within the discontinuities are usually allocated exclusively either to SLA or to commercial banks.

5. In 1955 only 9 per cent of all savers at SLA and (commercial and savings) banks split their accounts between the two types of institutions. Cf. *Savings and Loan News*, May 1956, p. 37.

stressed that the slopes of the indifference curves are not likely to be the same for all levels of savings.⁶ For any given individual, the marginal utility of the savings account dollar is presumably smaller for a large than for a small savings account. Accordingly, the slope is smaller for a large than for a small savings account, because the marginal utility of the savings account dollar would be greater for the latter. This is most important, because it means that, for the same relative rates of return, some individuals would shift the placement decision for savings funds, as size of savings account increases. This is consistent with the savings survey which showed that only 19 per cent of savers with family incomes of less than \$4,000 a year used SLA, whereas 27 per cent of savers in the \$4,000–\$6,000 income bracket, and 28 per cent of savers with more than \$6,000 incomes used SLA.⁷

Preference patterns and market shares. During the period studied, the commercial bank share of savings accounts declined. We have already demonstrated that the explanation for this shift in terms of interest rate differences is untenable. In the popular view, the explanation for the shift must thus fall exclusively on the increase in salesmanship. We have characterized the last position as seriously incomplete, because we believe the explanation for the erosion of the commercial bank share of the savings account market must include other important developments, like the growth in the number of savers, the changing size distribution of savers, and the changing preference patterns of savers for share accounts and time deposits. This conclusion follows from the indifference analysis and can be summarized in the following steps. First, the preference function of individual savers has changed. Although the slopes of the rate of return curves decreased, the slopes of the indifference curves decreased still more. This relative change induced some small savers to shift from time deposits to share accounts. For similar reasons, large savers have also shifted accounts. The change in preference functions and the corresponding shift of deposits was by no means exclusively related to SLA salesmanship. Partly, the change in preference patterns was related to the tremendous cushion of liquid assets accumulated by the public during the war and postwar years. These liquid assets fostered a greater sense of economic security, which, in turn, altered

6. We can ignore the fact that accounts over \$10,000 are not insured, because few savings accounts are likely to exceed that amount. For example, in September 1955, 99 per cent of the savings and time accounts of individuals, partnerships, and corporations at insured commercial banks were for \$10,000 or less. Cf. FDIC, *Annual Report*, 1955, p. 72.

7. Cf. *Savings and Loan News*, May 1956, p. 37.

attitudes towards risk-bearing in favor of higher interest incomes. Partly, too, these changing attitudes were re-enforced by rising income levels and the general atmosphere of prosperity during the postwar years.

Second, savings shares have altered because the size of the average savings accounts of individuals grew at both institutions, as income levels increased during the period studied.⁸ It will be recalled that, with no change in either interest rates or salesmanship — and, hence, no favorable alteration of the preference function — mere growth of an individual's savings account is sufficient, given the different slopes of the indifference curve for different savings levels, to shift some savings from commercial banks to SLA. In other words, a savings shift would have occurred with no change whatever in SLA salesmanship. In fact, SLA salesmanship was stepped up and was an *additional* factor in shifting savings deposits, despite the narrowing gap between interest rates. Here, again, the factors which were at work, quite independently of salesmanship, to change the subjective estimate and/or the disutility of risk at high savings levels were re-enforced by SLA advertising.

Third, the total number of savers increased during the period studied. It seems highly probable that many savers entered the scene with more favorable attitudes towards the riskiness of share accounts than was generally true of savers at the beginning of the period. Memories of SLA loss experience during the depression had dimmed and were probably unknown to some of the new savers whose placement decisions also contributed to the decline of the commercial banks' position.⁹ Here, again, increased salesmanship might have helped the dimming process for those who still remembered — though sheer passage of time is pretty effective by itself — but was hardly necessary to help those new savers forget who had no memories of the earlier losses.

Fourth, the postwar period has been characterized by a strong demand for housing. Inevitably, many mortgage seekers (especially for conventional loans) have found their way to SLA. This contact brought the SLA within the "orbit of comparison" of many borrowers,

8. Between 1948 and 1955, for example, the number of SLA share accounts increased from 8.9 million to 18.7 million. (Cf. *Savings and Loan Fact Book*, 1956, p. 52.) As we showed earlier, the volume of share accounts increased from \$11 billion in 1948 to more than \$32 billion in 1955. Thus, the average size of share account has also increased. Similarly, the average size of savings accounts at commercial banks has increased. (Cf. FDIC, *Annual Report*, 1955, p. 56.)

9. It should also be mentioned that between 1948 and 1954, the per cent of share accounts in insured SLA increased from 75.3 to 89.8 per cent. Cf. *Savings and Loan News*, Mar. 1956, p. 85.

who, in another capacity, were often also savers. The preference functions for at least some of these savers, given the relative rates of return, led them to save at SLA rather than at commercial banks. Again it seems certain that many mortgage seekers would have discovered the SLA without the aid of any increase in SLA salesmanship.

Finally, in our concern to press the many factors other than increased SLA salesmanship, we must not go to the extreme of denying the role played by salesmanship. We have tried to show that many factors operated to change the savings shares of SLA and commercial banks, independently of any increased salesmanship by SLA. It should not be ignored, however, that the SLA were not passive during the years under study. It has been estimated that in 1954 the SLA spent \$.97 per \$1,000 of share accounts on savings advertising, whereas the banks spent only \$.24 per \$1,000 of time deposits.¹ This advertising has been aimed at bringing SLA within the orbit of comparison of commercial bank and other savers, and also at altering the preference pattern in favor of the SLA package, and it undoubtedly re-enforced the factors which were already at work.

A cumulative process. The possibility of a cumulative relative decline of commercial bank savings is implicit in the foregoing. This depends upon the reaction of certain key variables to initial transfers of funds. A net transfer of funds from commercial banks to SLA may increase either the money supply (demand deposits and currency) or the velocity of money. The quantity of money will increase after a net transfer of time deposits to share accounts, because this increase in share accounts normally increases demand deposits, although not by the full amount of the decrease in time deposits. The velocity of money will increase after inactive demand deposits are transferred to share accounts. This transfer leaves the quantity of money unchanged, but SLA lending operations increase its velocity. In both cases, *ceteris paribus*, liquid assets of individuals increase.² Liquid assets increase after a net transfer of time deposits to share accounts, because, although the decline of time deposits is exactly balanced by an increase in share accounts, demand deposits increase. Liquid assets will also increase after inactive demand deposits are transferred to share accounts, because, although demand deposits are unchanged, share accounts increase. Finally, the increase in the quantity of money, following a transfer of time deposits to share

1. Cf. Freeman, *op. cit.*, p. 16. The figure for bank advertising on time deposits was calculated from Freeman's data.

2. Liquid assets include demand deposits, time deposits, and SLA share accounts. Cf. *Federal Reserve Bulletin*, July 1955, p. 749.

accounts, will normally increase the level of income, because SLA mortgage loans are now typically almost 100 per cent of share accounts. Similarly, the rise in velocity, after inactive demand deposits are transferred to share accounts, will also increase the level of income, because the money supply is unchanged.

As we have shown earlier, an increase in income or in liquid assets will tend to decrease the commercial bank share of savings, so long as the percentage of savings in savings accounts does not decline as income increases. To repeat, this is because the increase in both income and liquid assets tends to alter attitudes towards risk-bearing. Similarly, the increase in income will likely increase the average size of savings account. Both changes tend to diminish the commercial bank share of savings accounts. Under the circumstances we have described, an initial increase in share accounts may set in motion a cumulative process which results in a still further decline in the commercial bank share of savings accounts.

III. SOURCES OF COMPARATIVE COMPETITIVE STRENGTH

The erosion of the commercial bank position in the savings field is an accomplished fact. The interesting question is whether the position of these banks will deteriorate further or whether they can arrest and even reverse the present trend. The answer to this important question depends on the comparative competitive strength of commercial banks vis-à-vis their arch rivals for savings accounts, the SLA. Broadly speaking, the commercial banks can proceed along three main lines: increased product competition, increased price competition, and increased political activity. An analysis of increased political activity, e.g., to secure passage of legislation affecting commercial bank and SLA tax treatment, entry, control of lending ability, etc. — is beyond the scope of this paper.³ We here concentrate on the possibilities of increased price and product competition.

In the minds of many savers today, commercial banks and SLA are much more nearly on a par than has been true in the past. Effective product competition can substitute somewhat for price competition. For example, banks can make greater efforts to distinguish FDIC coverage from FSLIC coverage. Or banks can stress their superior liquidity, etc. Aggressive salesmanship along these lines might retard the rate of commercial bank decline in the savings field, but barring a general economic depression it is not likely to recapture the savings market for commercial banks from the by now well estab-

3. However, the economic impact of much proposed legislation can be more readily perceived by the analysis which follows.

lished SLA. Barring a general economic decline or special factors adverse to SLA, the struggle for the public's savings deposits must progressively be fought in the price sector.⁴ To assess the strength of commercial banks and SLA in this area involves an analysis of four major factors: (1) comparative costs, (2) comparative demand, (3) comparative use of savings deposits, and (4) comparative tax treatment.

Comparative cost factors. Dividends on share accounts have a dual aspect. On the one hand, dividends are a return to savers and directly influence SLA growth in terms of share accounts. On the other hand, dividends are a production cost and influence the SLA competitive position in the product market. The product market concerns us, because prices paid in the factor market must be compatible with earnings in the product market.

A word of explanation is in order for treating SLA dividends like interest payments. Interest on time deposits is unambiguously a production cost for banks, and the interest payment is a return to savers. By contrast, SLA dividends are a return to investors, and, as such, should not be treated like interest but like a distributive residuum. It is our contention that, despite the technical distinction between dividends and interest, they need not be treated differently in this paper. The reasons for our position are as follows: First, savers generally regard share accounts and time deposit accounts as equivalent. In all probability, many savers are not even aware of the technical distinction. For most savers, share account dividends and time deposit interest are both interest returns on savings. Second, SLA set dividend rates to compete with time deposit rates, not commercial bank dividend rates. Usually, too, SLA give six months' advance notice of changes in dividend rates. Third, the Internal Revenue Bureau permits SLA to deduct dividend payments as an interest cost before calculating taxes. Fourth, for tax purposes, savers must treat SLA dividends like interest returns, not dividends. For the cost analysis of this paper, therefore, SLA dividends are treated like commercial bank interest on savings accounts.

The innovation hypothesis. To treat dividends like interest and to include them as a production cost for SLA erects a hurdle in the product market which SLA must overcome to compete effectively with commercial banks. As we have shown, the effective interest

4. The price competition has already begun. During the period of our study, there was no significant pressure to raise the ceiling rate on member bank time deposits. Late in 1956, however, the ceiling rate was raised, and many commercial banks are now engaging SLA in more vigorous rate competition.

rate has been more than twice as high for SLA as for commercial banks. SLA growth must hurdle higher dividend costs, and the interesting question is how they can do it. One possible answer is that SLA enjoy more than compensating cost advantages in other directions, which permit them to compete successfully with commercial banks on the product market. The most obvious cost advantage inheres in the mutual organization of SLA. Mutual associations need not return profits on equity investment as banks must do. To be sure, no profit-making enterprise need earn profits in every period; indeed, profits are not an operating cost, but a residue from operations. While this is undoubtedly correct for any single period, the position is untenable over several periods. Over a period of years, investors will demand at least a minimum normal return or they will understandably seek to liquidate their holdings for more profitable reinvestment elsewhere. In other words, a certain minimum profit is part of the long-run supply price of invested capital. Hence, it is legitimate to maintain that such profits are part of the long-run cost of production. Since mutuals are spared this production cost, this is a source of competitive advantage in the product market.

The preceding argument amounts to saying that SLA are innovating firms. The nature of their innovation is a new form or organization, viz., the mutual form.⁵ A hallmark of innovation is cost reduction which permits the new firm to outcompete the old, "obsolete" firms at prevailing product price levels, and, often, to outsurvive older firms in the face of falling price levels.

The novelty of the mutual form of organization among financial intermediaries is not denied. The question is whether this novelty qualifies as a bona fide innovation. In a manufacturing firm, an innovation can be detected by a change in the production function, or, alternatively, by a change in the long-run cost function which reveals explicitly the relation between the productivity and factor price elements inherent in the scale line.⁶ The long-run cost function for financial intermediaries can be approximated by plotting average costs (total costs divided by volume of loans and investments) for different size firms. In recent years, this would show a declining cost pattern for both institutions as size of firm increases. Moreover, even the smallest category of commercial banks would show lower unit costs than the largest category of SLA.⁷ Unfortunately,

5. SLA are not really new, but they have competed effectively with banks only since their reorganization during the thirties.

6. Cf. J. A. Schumpeter, *Business Cycles*, I, 87-89.

7. Commercial bank costs by size of bank can be calculated from figures in FDIC, *Annual Report*; SLA costs, from the Federal Home Loan Bank Board, *Combined Financial Statements*.

our test is not conclusive, for, although SLA are overwhelmingly single-product firms, commercial banks are multiproduct firms. Hence, it is necessary to isolate mortgage costs from the production costs of other bank products.

It might as well be admitted at the outset that this cannot be accomplished to complete satisfaction with existing data. Although definite results are not possible, highly suggestive results are possible. Let us begin with the direct money costs for both institutions. In 1954, SLA paid interest-dividends at the rate of 2.88 per cent; commercial banks, at the rate of 1.32 per cent. (See Table I.) To make the money costs more nearly comparable, we add the prorated share of profit (dividends on invested bank capital) cost to the time deposit cost. This amounts to \$.28 per \$100 of time deposits, and raises the money cost on time deposits from 1.32 per cent to 1.60 per cent.⁸ The average costs (total costs divided by loans and investments) of SLA in 1954 were 4.07 per cent.⁹ Since direct money costs were 2.88 per cent, operating costs averaged about 1.2 per cent for all SLA.¹ In 1954 the difference between SLA and commercial bank money costs was 1.28 per cent (2.88 less 1.60), or almost identical with the operating costs of SLA. This leaves a margin of 2.47 per cent (4.07 less 1.60) between the mortgage money costs of commercial banks and the full costs of SLA. In other words, for SLA to have an innovating cost advantage, commercial bank operating costs on mortgages alone would have to be more than 2.47 per cent.

To repeat, we cannot get direct operating costs on mortgages for commercial banks, but such information seems hardly necessary

8. Computed from FDIC, *Annual Report*, 1954, pp. 146-47. This calculation is in keeping with our earlier decision to treat profits as a production cost. The rationale is as follows: SLA do not pay profit-dividends; commercial banks do. In a sense, therefore, the profit-dividends are legitimately part of the commercial banks' cost of funds — a cost not incurred by the SLA. To make the cost figures for both institutions as nearly comparable as possible, we have prorated the commercial bank dividend payments to their deposits, both demand and time. Thus, implicitly, we are saying that the cost of each dollar of borrowed funds (whether demand or time) must be increased by its proportionate share of the dividend payments. Some might even argue that the banks' capital funds should be included in the total funds base before allocating profit costs. By omitting equity funds from the base, we have taken the less, not the more conservative position.

9. Computed from Federal Home Loan Bank Board, *Combined Financial Statements*, 1954, pp. 12, 48-49. Since there was a considerable growth in loans and investments during the year, we averaged the volume of loans and investments as of Dec. 31, 1953, and 1954, for the denominator of the ratio.

1. Norman Strunk estimated that operating costs for a home mortgage business are, at a minimum, 1 per cent. See his chapter, "Urban Real Estate Financing," in Herbert V. Prochnow (ed.), *American Financial Institutions*, p. 158.

to resolve the point at issue. For commercial banks to be higher cost producers of mortgages, commercial bank operating costs would have to be, at a minimum, slightly more than twice as high as SLA operating costs and that is surely highly improbable.² No obvious reason suggests itself for such terrific "inefficiency" by commercial banks. We conclude that it is highly probable that SLA have no cost advantage over commercial banks in producing mortgage loans. On the contrary, SLA appear to have a cost disadvantage.

Multiple products vs. specialized firm hypothesis. We have shown that SLA and commercial bank competition in the product market is not explained by the innovation hypothesis. The facts in this case seem rather to fit a different hypothesis, viz., that SLA are specialized firms (in mortgages) competing with one of the products (mortgages) of multiproduct firms (commercial banks) in which the cost advantage is with the banks. The theory of multiproduct firms tells us that the opportunity for multiple products arises when a firm has excess capacity which can be absorbed by the new product at a lower cost than would be incurred by a new company set up exclusively for that product.³ The multiple products hypothesis thus has two elements: excess capacity and a cost advantage. Commercial banks appear to qualify on both accounts.

First, commercial banks qualify by having excess capacity. The primary business of commercial banks is lending. Historically, at least, bank lending meant mostly short-term loans to business. Banks have always held a certain volume of nonloan assets for portfolio reasons (e.g., liquidity), but this did not confuse bankers about their primary responsibility in the credit field. During the secular decline of business loans from 1920-1921 to the end of World War II, bankers were plagued with excess capacity in the form of lending power in excess of the demand for suitable business loans. This excess capacity became chronic following the enormous expansion of the money supply during World War II. The resources of the banking system expanded far more than the demand for business credit. Over the years this excess capacity has motivated banks to extend credit in new (for them) fields, especially in those of consumer and real estate loans,⁴

2. Full costs would also include mortgage loan losses. At present, such losses are very small, and future losses on present holdings are extremely difficult to predict. If reserves for future loan losses are considered part of costs, the higher percentage of insured mortgage loans in banks' mortgage portfolios would certainly be consistent with lower reserves per dollar of mortgage loans for the commercial banks. In general, in this paper, we deal only with current yields and expenses, omitting possible long-run losses.

3. Cf. Joel Dean, *Managerial Economics*, pp. 115, *passim*.

4. This is not to deny the importance of the profitability and portfolio role of these new forms of credit.

as well as in bonds for income (as contrasted with bonds for liquidity). As a result, most commercial banks today are full-fledged multi-product firms, and most of the new products appear to be permanent. Second, commercial banks also qualify under the multiproducts hypothesis because of their cost advantage on mortgage loans. We have already demonstrated that commercial banks can probably produce mortgage loans more cheaply than SLA, even on comparative full costs. In practice, however, mortgage loans of a commercial bank need cover only marginal costs to be profitable for the bank.⁵ It seems reasonably certain, therefore, that the multiproduct commercial banks can produce mortgage loans at a lower cost than the highly specialized SLA.

The multiple products versus specialized firm hypothesis presented here to explain SLA and commercial bank competition in the product market has not been elsewhere disputed; it has been ignored. There appear to be three explanations for this fact. First, it has been common to attribute mortgage lending in commercial banks to time deposits rather than to excess capacity. This confusion probably originated during the period when the real bills doctrine was the dominant principle of portfolio composition. From the valid notion that time deposits could properly be invested in long-term paper so far as they represented savings,⁶ it was easy to infer the mistaken notion that time deposits were responsible for long-term paper in bank portfolios. For sound portfolio management, time deposits may indeed be necessary for mortgage lending, but time deposits per se do not make mortgage lending desirable.

Second, the multiple products hypothesis, with its stress on excess capacity, may also have been ignored because it seemed to be irrelevant. There is a prevailing notion that mortgage loans are desirable for their earnings. If this were the case, banks with time deposits would seek mortgage loans without reference to excess capacity. The fact is that mortgage loans are not independently desirable for their earnings, compared with the average returns on the balance of a bank's loan portfolio. To prove our contention, we need to compare the average return on mortgages with the average return on all other loans. This information is not directly available, but we can approximate it indirectly. The average return on mem-

5. The marginal decision for a bank embraces the category of mortgage loans, not just any single mortgage loan, though the latter is also involved.

6. Mints has cited the Federal Reserve position that "... banks carry time deposits and ... since these deposits represent savings, the banks may properly invest part of their funds in long-term paper." Lloyd Mints, *A History of Banking Theory*, p. 269.

ber bank total loans is higher than the average return on member SLA mortgages.⁷ It will be shown later that SLA earn more on mortgages than commercial banks. Therefore, member bank earnings on all except mortgage loans are higher than on mortgage loans. Moreover, average loan costs for mortgages are not less than average costs for all other loans combined. We conclude, therefore, that mortgage loans must earn less than the average return on the balance of the bank's loan portfolio.

Finally, it has been common to overstate the extent of competition in commercial banking, with a failure to be alerted to oligopoly strategy in banking markets. In fact, most banks operate in local loan markets with a high degree of market concentration. An oligopolistic price stabilization can degenerate into cutthroat price competition under the pressure of excess capacity. As Robert Lanzilotti has shown, the development of a new product to absorb excess capacity can help to forestall ruinous price competition.⁸

Our examination of comparative costs strongly suggests that commercial banks could afford to pay the same interest rates as SLA — certainly they could narrow the gap — if costs were the only relevant consideration. In fact, commercial banks do not pay the same rate, and we have shown this is not because of an inherent cost disadvantage. Since the explanation for the failure of commercial banks to pay higher interest rates cannot be discovered on the supply side of the market, we turn to an examination of demand in the product market.

Cross-elasticities of demand. The disparity between SLA and commercial bank interest stems partly from the nature of their product markets. For our purposes, the relevant product market can be subdivided into three categories: (1) insured residential mortgages, (2) noninsured residential mortgages, and (3) commercial and industrial mortgages. In this analysis, the relevant demand concept is the cross-elasticity of demand for the products of both lenders.

Cross-elasticities of demand are not the same for all mortgage submarkets. For example, in the market for insured residential mortgages, the cross-elasticity of demand approaches infinity for

7. Cf. Federal Home Loan Bank Board, *Combined Financial Statements*, 1954, p. 7, and *Federal Reserve Bulletin*, "Member Bank Operating Ratios," (averages of individual bank ratios) for each year. Generally, only the two largest size categories of member banks earned less on total loans than SLA earned on mortgages.

8. Cf. Robert Lanzilotti, "Multiple Products and Oligopoly Strategy: A Development of Chamberlin's Theory of Products," this *Journal*, LXVIII (Aug. 1954), 461-74.

both SLA and commercial banks.⁹ For SLA, the cross-elasticity of demand for noninsured residential (conventional) mortgages is high, but distinctly smaller than for insured residential mortgages. Moreover, the cross-elasticity of demand for noninsured residential mortgages is greater for commercial banks than for SLA. On the other hand, the cross-elasticity of demand for commercial and industrial mortgages is greater for SLA than for commercial banks, and is greater for both lenders than the cross-elasticity of demand for residential mortgages.

The reasons for the partial insulation of some mortgage submarkets and for the different cross-elasticities are partly legal and partly traditional. For example, the legal loan-to-value ratio on noninsured residential mortgages is higher for SLA than that permitted for commercial banks. Other things being equal, this would lower the cross-elasticity of demand for SLA for the category of borrowers interested in the nonoverlapping range of loan-to-value conventional mortgages. Similarly, the permissible maximum maturity of commercial bank mortgages (except insured mortgages) has been substantially less than for SLA. Other things being equal, this would tend to reduce commercial bank participation in the comparatively high-earning conventional mortgage field. In general, in the residential mortgage area commercial banks face a more restrictive legal environment than SLA. On the other hand, federal SLA "... are prohibited, with certain limited exceptions, from investing more than 15 per cent of their assets . . . on properties other than one-to-four family residences."¹ This prohibition restricts SLA participation in the commercial and industrial mortgage market.²

Commercial banks are bound by their traditional conservatism as well as by legal restrictions. This conservatism is particularly reflected in the heavy reliance on mortgage insurance by commercial banks. Moreover, commercial banks take only the preferred risks among conventional loans — and charge lower rates on such mortgages,³ whereas SLA are clearly disposed to assume riskier loans. These different attitudes to risk-bearing also segment the mortgage market.

The segmentation of the mortgage market and the different

9. In practice, this has been somewhat obscured, because FHA and VA mortgage rates are administratively determined, and the official rate is usually the actual rate.

1. Norman Strunk, *op. cit.*, pp. 174-75.

2. At the present time SLA are below their legal limit on this type of mortgage.

3. Cf. Roland I. Robinson, *The Management of Bank Funds*, p. 234.

cross-elasticities of demand in the various submarkets are reflected in the rates earned on different mortgages. For example, in 1947 both lenders earned higher average returns on conventional mortgages than on insured mortgages.⁴ Similarly, early in 1955 the most typical rate on conventional loans made by SLA was $4\frac{1}{2}$ per cent and under for only 2 per cent of SLA, 5 per cent for 46 per cent of SLA, $5\frac{1}{2}$ per cent for 16 per cent of SLA, and more than $5\frac{1}{2}$ per cent for fully 36 per cent of the reporting institutions.⁵ Thus, when more than half of the SLA were charging $5\frac{1}{2}$ per cent or more on conventional mortgages, the insured mortgage rate was only $4\frac{1}{2}$ per cent on FHA and VA mortgages. Not only are average rates on conventional mortgages higher than on insured mortgages, but SLA also earn higher rates on conventional mortgages than commercial banks.⁶ Because of the cross-elasticities, it is also not surprising that both lenders earn more on residential mortgages than on commercial and industrial mortgages.⁷

The significance of these rate differences for total mortgage earnings is more fully appreciated in connection with portfolio distributions for both lenders. In 1947, for example, 44 per cent of outstanding bank residential loans on one-to-four family homes earned less than 5 per cent, compared with 13 per cent for SLA; 16 per cent of the banks' volume carried a contract rate of 6 per cent or more, compared with 29 per cent for SLA.⁸ Or, again, at the end of 1955, 52 per cent of the total volume of commercial bank residential mortgages was insured; only 23 per cent of SLA mortgage volume was insured.⁹

We have shown that the mortgage market is not a homogeneous market and that cross-elasticities of demand are not the same for all submarkets of the mortgage market. The weight of the evidence further suggests that SLA concentrate more than commercial banks in the higher-yielding segments of the market. To this extent, SLA can pay higher interest rates than commercial banks.

Comparative use of savings deposits. The ability to engage in rate competition is also affected by the comparative use of savings

4. J. E. Morton, *Urban Mortgage Lending: Comparative Markets and Experience* (Princeton, 1956), p. 82.

5. Based on a survey reported by Don. M. Dailey in *Savings and Loan News*, June 1955, p. 71.

6. Morton, *op. cit.*, pp. 80, 81, 173.

7. *Ibid.*, p. 82.

8. *Ibid.*, p. 81.

9. Computed from *Federal Reserve Bulletin*, Nov. 1956, pp. 1225-26. For 1951 the corresponding figures were 50 per cent for commercial banks and 26 per cent for SLA.

deposits in SLA and commercial banks. As one commercial banker put it, SLA "... can and do pay more attractive rates of return because they earn more, through investing nearly 100 per cent of their savings in high-rate, long-term mortgage loans."¹ By contrast, commercial bank mortgages account for only 30 to 45 per cent of time deposits.² Moreover, commercial banks invest a substantial proportion of the balance of their time deposits (other than legal reserves) in bonds, and bank investments generally earn less than mortgage loans.

The low ratio of mortgages to time deposits in commercial banks is partly owing to an absence of suitable mortgage applications. Similarly, the high bond investment ratios are partly a residual employment of excess capacity for which no suitable loan alternatives (of all kinds) are available. Banker conservatism about liquidity also explains low mortgage ratios. Mortgages are long-term investments and bankers want them to be buttressed by ample liquidity. Savings account turnover is an important determinant of liquidity need, and bank turnover is much higher than SLA turnover.³ Moreover, during the last twenty years SLA have been willing to rely heavily on the Federal Home Loan Banks to meet unusually severe withdrawals, whereas commercial banks are often reluctant to borrow at the Federal Reserve. Finally, in past years SLA managers worried less than commercial bankers about liquidity, because SLA were not considered to be demand institutions. At present, however, SLA must operate on what amounts to a demand basis. Accordingly, the liquidity problem is receiving more attention among SLA managers than ever before.⁴

Comparative tax treatment. One final factor may influence the comparative interest payments on savings accounts by commercial banks and SLA — their different tax treatment. Since 1952 both have been subject to federal income taxes. Both also set aside reserves for losses. However, the tax treatment for SLA losses is generally conceded to be more generous at present than that for

1. Freeman, *op. cit.*, p. 27.

2. The range is for different size banks. Calculated from figures in FDIC, *Annual Report*, 1954, pp. 132-33 and 134-35.

3. In 1953, total savings account withdrawals from banks were 46 per cent of the average amount of time deposits, but only 27 per cent of average savings at SLA. Withdrawals relative to new deposits also appear to be higher for banks than for SLA. Cf. *Banking*, Aug. 1955, p. 48, and *Savings and Loan News*, Jan. 1955, p. 69. Cf. also U. S. Savings and Loan League, "Quarterly Letter on Savings and Home Mortgages," Sept. 1951.

4. Cf. for example, *Savings and Loan News*, June 1955, p. 36, and Dec. 1955, pp. 39-40.

commercial banks. Clearly, differential tax treatment is not an element of economic efficiency, but equally clearly, differential tax treatment can influence comparative interest payments on savings accounts.

In summary, commercial banks are at no competitive disadvantage with SLA in rate competition for savings so far as costs are concerned. They are at a disadvantage, however, because of their lower earnings on mortgages, a smaller percentage of savings deposits in mortgages, and, possibly, the preferential tax treatment for SLA.

IV. CONCLUSIONS

1. The erosion of the formerly dominant position of commercial banks in the savings field has been concomitant with the growing institutionalization of the savings process. During the postwar period the most serious loss of position has been to SLA. An understanding of the forces responsible for this shift among financial intermediaries is important both for theory and for public policy.

2. The generally accepted explanation for the shift in position is the interest rate difference coupled with the increased salesmanship of SLA. This popular explanation is superficial; it is half wrong; and it is seriously incomplete in the remainder. It is superficial because it leaves many important questions unanswered; it is half wrong because the interest rate difference has actually declined (or remained constant in absolute terms) precisely during the period of rapid relative growth of SLA; it is seriously incomplete in the remainder because, during the period studied, many factors were involved in the relative decline of commercial bank savings, quite independently of increased salesmanship by SLA. The major factors responsible for the shift have been the increased size of savers, the increase in the total number of savers, the construction boom, and the changed preference functions of individual savers in favor of SLA. The changed preference patterns are, in turn, related to the growth of income, the growth of personal savings, the increase in liquid assets of individuals, and, as a final re-enforcing factor, the intensive product competition by SLA. The relative shift was probably also re-enforced by a cumulative process engendered by the initial transfers.

3. The future struggle for savings accounts must concentrate on price and product competition and political action. Barring a serious economic decline or special circumstances adverse to SLA, product competition probably cannot reverse the present trend, though it can retard its rate. Any significant reversal of present shares would have to involve price (rate) competition.

4. The probable outcome of price competition depends on the comparative competitive strength of commercial banks and SLA. In turn, this depends on comparative cost factors, comparative demand, comparative use of savings accounts, and comparative tax treatment. The commercial bank savings deposit rate need not be lower than SLA rates because of any cost handicap of the former. Commercial bank rate competition is hampered, however, by lower earnings on mortgages, a smaller percentage of savings deposits in mortgages, and, possibly, preferential tax treatment for SLA. Other things being equal, therefore, more effective rate competition by banks will probably impair their net earnings position.

DAVID A. ALHADEFF.

UNIVERSITY OF CALIFORNIA, BERKELEY.

CHARLOTTE P. ALHADEFF.

BERKELEY, CALIFORNIA.

FORECASTING USES OF ANTICIPATORY DATA ON INVESTMENT AND SALES*

By F. MODIGLIANI AND H. M. WEINGARTNER

I. Introduction, 23; purpose of the analysis, 23; nature of the data, 25. — II. Direct forecasting value of the anticipatory data, 26; capital outlay plans, 26; sales anticipations, 34. — III. Errors of sales anticipations and the realization of investment plans — an example of the realization function approach, 36; statement of the problem, 36; formulation of the model, 39; empirical results, 43; some analytical implications, 47; some forecasting implications, 50. — Appendix, 53.

I. INTRODUCTION

A. *Purpose of the Analysis*

The systematic collection of data on expectations and plans of firms dates back almost two decades and has grown in importance since the end of World War II.¹ The purpose of the present paper is to explore the usefulness of survey data on capital outlay plans and sales expectations in economic forecasting. Of the three main series on investment plans which are currently being compiled in the United States, this paper focuses on only one, namely, the "Annual Survey of Business Anticipations of Plant and Equipment Expenditures" collected by the Securities and Exchange Commission and the Department of Commerce.²

Several previous investigations of this and other surveys have concentrated on cross-section analyses, i.e., on the relation between

* The present paper is an outgrowth of the project "Expectations and Economic Fluctuations" financed by a grant from the Merrill Foundation for the Advancement of Financial Knowledge, and carried out at the University of Illinois under the direction of the senior author. The research reported herein was supported in part by a research grant from the Graduate School of Industrial Administration, Carnegie Institute of Technology. The paper was completed at Carnegie Institute of Technology in May, 1957.

1. Probably the earliest efforts in this area are represented by the survey of capital outlay plans carried out in Sweden since 1930. Since that time the collection of similar data has been imitated in many other countries, including the United States. In the postwar period surveys have also been conducted on a number of other anticipatory variables, such as prices, profits, sales, and employment. Notable examples of such surveys are those conducted by Dun and Bradstreet in the United States, and by the IFO-Institute in Germany.

2. The other surveys are the McGraw-Hill Publishing Company's Department of Economic Surveys, "Business' Plans for New Plants and Equipment"; and the "Quarterly Survey of Capital Appropriations of the 1000 Largest Manufacturing Firms," conducted by the National Industrial Conference Board and published in *The Conference Board Record*.

anticipations and realizations of individual firms. The present paper, on the other hand, while it has benefited from these earlier studies, notably those of Friend and Bronfenbrenner and of Eisner,³ is concerned exclusively with the time-series analysis of the data, that is, with assessing the usefulness of aggregate investment plans and sales anticipations in forecasting broad aggregates. Such a time-series analysis is the most relevant one in so far as one is interested in forecasting aggregates; cross-section analyses are of only indirect value in this respect.

However, up to this date a series of sufficient length for time-series analysis has not been available. At this writing the annual SEC-Commerce Survey provides data covering both investment plans and sales expectations for a period of nine years. This is a body of data which is by now worth exploring, even though conclusions derived from it must be regarded as somewhat tentative.

Section II of this paper is devoted to an analysis of the record of the anticipatory data as direct forecasts. Simple comparisons between anticipated and actual investment have been made previously by a number of authors, especially in the *Survey of Current Business*.⁴ The analysis presented in Section IIA differs from these earlier studies in that it attempts to evaluate systematically the predictive ability of investment plans by developing various relevant standards of comparison. Furthermore, the desirability of making certain adjustments to the data, notably adjustments for changes in price level, is examined and tested. In Section IIB a similar analysis is applied to the sales anticipations reported by the Survey, a body of data which, so far, has received little attention.

Section III analyzes possible uses of the anticipatory data other than in direct forecasts. Here primary reliance is placed on the "Realization Function Approach," a theoretical framework which relates the realization of investment plans to the realization of sales and other anticipations. A brief outline of this framework is provided

3. I. Friend, and J. Bronfenbrenner, "Business Investment Programs and Their Realization," *Survey of Current Business*, Dec. 1950, and "Plant and Equipment Programs and Their Realization," *Short-Term Economic Forecasting (Studies in Income and Wealth, XVII, National Bureau of Economic Research, Princeton, 1955)*, pp. 53-111; and R. Eisner, "Expectations, Plans and Capital Expenditures; A Synthesis of *Ex Post* and *Ex Ante* Data (Based on the McGraw-Hill Data)," *Proceedings of the Conference on Expectations, Uncertainty, and Business Behavior*, Social Science Research Council, October 27-29, 1955, forthcoming.

4. Cf. also "Statistics on Business Plant and Equipment Expenditure Expectations," *Report of the Consultant Committee on Business Plant and Equipment Expenditure Expectations*, organized by the Board of Governors of the Federal Reserve System at the request of the Sub-Committee on Economic Statistics of the Joint Committee on the Economic Report, Washington, D.C., July 1955, especially p. 32.

in Section IIIA. This is followed in Section IIIB by the formulation of a simple model which demonstrates this approach, and which is capable of empirical testing with the data at hand. The empirical results of the tests, which were generally quite satisfactory, are reported in Section IIIC, while additional consequences of the model are discussed in the following section. Finally, some forecasting implications of this approach are presented in Section IIIE.

B. Nature of the Data

The Commerce-SEC data used in the analysis are estimates of national aggregates of plant and equipment expenditures of the previous year, and anticipated expenditures for the current year, based on questionnaires returned by a sample of firms in each of the industrial categories. Because the anticipations and actual expenditures are reported in different years, problems of comparability arise on several levels. First, at the level of the firm, because of personnel turnover and other reasons, changes in interpretation and reporting methods may occur between the time the anticipation is filed and the time the actual expenditures are reported. Second, the sample of respondents also changes somewhat from year to year, as does the universe of firms in the various industries. These changes must be recognized in the method of "blowing up" the sample to obtain an estimate of the aggregate statistic. The method used has actually been altered several times since the series was begun in 1947. In fact, in several instances the procedure employed in estimating anticipated expenditures from the sample at the beginning of the year was different from that used in estimating actual expenditures at the end of the year.

For these reasons published figures on anticipated expenditures are not directly comparable with the latest and presumably most reliable figures for actual expenditures. As a result the data on anticipated expenditures presented in the table below involve certain adjustments and do not necessarily agree with the latest published figures.⁵ The purpose of these adjustments is to make the series as far as possible comparable with the latest Commerce-SEC series on actual outlays.⁶

5. The figures on revised anticipated capital outlays can, however, be reconstructed from the table of percentage changes in anticipated and realized investment given in M. Foss and V. Natrella, "Ten Years' Experience With Business Investment Anticipations," *Survey of Current Business*, Jan. 1957, p. 17. Unrounded figures were made available to the authors through the co-operation of Mr. Larry Bridge of the Department of Commerce.

6. It should be noted in this connection that what one would really like to test is whether the anticipations accurately forecast *actual* investment. Operationally, however, the only thing it is possible to test is how well they forecast the Commerce-SEC *estimates* of actual investment.

II. DIRECT FORECASTING VALUE OF THE ANTICIPATORY DATA

A. Capital Outlay Plans

The first three columns of Table I present for each year actual expenditures, I_t , expenditures anticipated at the beginning of the year, $I_t(t-1)$ and the forecast error as a percentage of actual expenditure. Part A of the table relates to manufacturing and Part B to the aggregate of all industries included in the survey. As an indication of over-all forecasting accuracy of the anticipatory data, the average absolute error of forecast is presented in row $a(i)$ of column (3). This measure is the average error computed by disregarding signs, so that errors of opposite sign do not offset each other.

The average error, as well as the errors for individual years, appears to be relatively small with the single exception of 1950. The large error in this year is not surprising, however, because the beginning of the Korean War occurred in the middle of that year and its consequent effects on investment activity could hardly have been anticipated at the beginning of the year. It appears therefore that the omission of the year 1950 may serve to provide a more reliable indication of the performance of the anticipatory data. Accordingly all measures of accuracy presented in the table have been computed including and excluding 1950. The significance of the figures of column (3) is difficult to evaluate in the absence of an appropriate yardstick. One useful standard which has been frequently utilized in recent years in assessing the quality of forecasts is represented by the error resulting from the so-called "naive model": "next year will be like this year." Denoting by I_t the level of investment expenditure in any year t , the naive model forecast for the year t is simply I_{t-1}

and the error of such a forecast is given by the formula $\frac{I_{t-1} - I_t}{I_t}$,

which has been used in computing the entries of column (4).

There are two distinct, though closely related, senses in which the naive model error of column (4) can be regarded as a yardstick for the forecast error of column (3).⁷ In the first place, a naive model forecast is the cheapest and simplest method of forecasting one can conceive of. It represents therefore a sort of minimal standard, a zero point, from which to measure the performance of any alternative forecasting procedure. Unless the alternative method can perform at least as well, it is essentially worthless. Alternatively, one may

7. Cf. F. Modigliani and O. H. Sauerlander, "Economic Expectations and Plans of Firms in Relation to Short-Term Forecasting," *Short-Term Economic Forecasting*, XVII, *op. cit.*

note that the naive model error for any given year is, by definition, identical with the yearly movement of the series — except for sign. Hence the average absolute error, shown in rows *a*(i) and (ii) of column (4), represents the average year to year variation and may be regarded as a measure of the difficulty inherent in forecasting the variable under consideration.

It appears from column (4) that, in spite of the stability of the postwar period, investment expenditure has been subject to considerable yearly fluctuations. When this variability is taken into account, the forecasting record of the anticipatory data appears quite favorable, as can be seen by comparing columns (3) and (4). In particular, if 1950 is omitted, the naive model error turns out to be nearly five times higher than that of the anticipatory data in the case of manufacturing, and nearly three times higher for the aggregate of all industries covered by the survey.⁸ This same result can be usefully expressed in terms of the "coefficient of relative accuracy" computed by subtracting the average forecast error of column (3) from the average error of the naive model in column (4), and expressing the difference as a percentage of the naive model error. This coefficient, shown in rows *b*(i) and (ii) of the table, measures the proportion of the variability of the series which has been correctly forecasted. It has a maximum value of 100 if the forecast is perfect throughout, a value of zero when the forecast is on the average no better than a naive forecast, and a negative value when it is worse.

It may be noted that while in terms of the average absolute error the forecasting performance of the manufacturing sector appears somewhat poorer than that of all industries, the difference very nearly vanishes in terms of the coefficient of relative accuracy. The larger absolute error of the manufacturing sector appears to be accounted for by the greater short-run variability, which in turn is likely to reflect, at least in part, the general tendency for individual components to fluctuate more than broad aggregates.

A test based on the naive model which has been used so far may be regarded as too lenient. Since the investment series is available on a quarterly basis, one could base the naive model forecast on the latest (seasonally adjusted) quarterly rate, rather than on the average rate of the last four quarters as is done in the test of column (4).

The result of using the fourth quarter of a year to forecast the following year's expenditure on plant and equipment is shown as

8. The figures of column (4) probably tend to overestimate the success of naive model forecasts since they are based on the latest revised figures rather than on the preliminary estimates actually available at the relevant points of time.

TABLE I
PLANNED AND ACTUAL CAPITAL OUTLAYS, AND SELECTED MEASURES OF FORECAST ACCURACY
A — MANUFACTURING

Year	Actual Expenditure (I_t) (1)	Planned Expenditure ($I_t(u-1)$) (2)	Raw Data		Forecast Error (%)			
			Investment Plans ¹ (3)	Naive Model I ² (4)	Naive Model II ³ (5)	Naive Model I ⁴ (6)	Naive Model II ⁵ (7)	Investment Plans ⁶ (8)
1 — 1947	8,703			-4.7	-1.4	3.9	4.0	-1.9
2 — 1948	9,134	8,499	-7.0	27.8	24.2	32.3	23.3	-0.8
3 — 1949	7,149	7,139	-0.1	-4.6	-14.8	-2.6	-12.5	-15.7
4 — 1950	7,491	6,143	-18.0	-31.0	-17.8	-25.0	-13.9	4.9
5 — 1951	10,852	10,863	0.1	-6.7	0.5	-5.6	1.2	1.9
6 — 1952	11,632	11,769	1.2	-2.3	-1.5	-0.2	0.2	-4.6
7 — 1953	11,908	11,675	-2.0	7.9	7.2	8.4	6.9	0
8 — 1954	11,038	11,068	0.3	-3.5	-7.5	-1.9	-5.8	-0.3
9 — 1955	11,439	10,704	-6.5					
Summary Measures of Forecast Accuracy								
a — Average Absolute Error								
(i) All years			4.4	11.1	9.4	1.00	8.5	3.8
(ii) Excluding 1950			2.5	12.0	8.6	11.0	7.9	2.1
b — Coefficient of Relative Accuracy								
(i) All years				60 ⁷	53 ⁷	62 ⁹	55 ⁹	
(ii) Excluding 1950				79 ⁸	71 ⁸	81 ¹⁰	73 ¹	
c — Arithmetic Mean Error								
(i) All years			-4.0	-2.1	-1.4	1.2	0.4	-2.1
(ii) Excluding 1950			-2.0	-1.8	0.5	1.7	2.3	-0.1

For footnotes, see Table B.

B — ALL INDUSTRIES

Year	Actual Expenditure (I_t) (1)	Planned Expenditure ($I_t(t-1)$) (2)	Forecast Error (%)					
			Raw Data	After Adjustment for Changes in Prices				
			Investment Plans ¹ (3)	Naive Model I ² (4)	Naive Model II ³ (5)	Naive Model I ⁴ (6)	Naive Model II ⁵ (7)	Investment Plans ⁶ (8)
1 — 1947	20,612		-2.7	-6.6	-3.4	1.8	1.8	2.5
2 — 1948	22,079	21,470	1.0	14.5	15.4	18.6	14.6	0.3
3 — 1949	19,285	19,487	-13.3	-6.4	-13.6	-4.5	-11.2	-10.9
4 — 1950	20,605	17,863	1.8	-19.6	-9.1	-12.7	-4.8	6.7
5 — 1951	25,644	26,114	0.1	-3.2	0.3	-2.0	1.0	0.8
6 — 1952	26,493	26,519	-4.5	-6.5	-5.7	-4.5	-4.0	-2.8
7 — 1953	28,322	27,051	1.3	5.6	6.3	6.1	6.1	1.1
8 — 1954	26,827	27,180	-5.7	-6.5	-8.9	-4.9	-7.1	-4.0
9 — 1955	28,701	27,063						
Summary Measures of Forecast Accuracy								
a — Average Absolute Error			3.8	8.6	7.8	6.9	6.3	3.6
(i) All years			2.4	8.9	7.0	7.2	5.6	2.6
(ii) Excluding 1950								
b — Coefficient of Relative Accuracy								
(i) All years				56 ⁷	51 ⁷	48 ⁸	43 ⁹	
(ii) Excluding 1950				73 ⁸	66 ⁸	64 ¹⁰	54 ¹⁰	
c — Arithmetic Mean Error			-2.8	-3.6	-2.3	-0.3	-0.5	-0.8
(i) All years			-1.2	-3.2	-0.7	0.3	1.1	0.7
(ii) Excluding 1950								
<hr/>								
1. $\frac{I_t(t-1) - I_t}{I_t} \times 100.$	2. $\frac{I_t - I_t}{I_t} \times 100.$	3. $\frac{I_t^{IV} - I_t}{I_t} \times 100.$	4. $\frac{I_t - I_t}{I_t} \times 100.$	5. $\frac{I_t^{IV} - I_t}{I_t} \times 100.$	6. $\frac{I_t(t-1) - I_t}{I_t} \times 100.$	$\frac{I_t^{IV} - I_t}{I_t} \times 100.$		

$$1. \frac{I_t(t-1) - I_t}{I_t} \times 100. \quad 2. \frac{I_{t-1} - I_t}{I_t} \times 100. \quad 3. \frac{I_{t-1} - I_t}{I_t} \times 100. \quad 4. \frac{I_{t-1} - I_t}{I_t} \times 100. \quad 5. \frac{I_{t-1} - I_t}{I_t} \times 100. \quad 6. \frac{I_{t-1} - I_t}{I_t} \times 100.$$

7. Difference between naive model error (row a (i)) and error of anticipations (Col. (3), row a (i)), expressed as percentage of naive model error.

8. Same as in note 7, but with row a (i) replaced by row a (ii).

9. Difference between naive model error (row a (i)) and error of anticipations adjusted for price changes (Col. (8), row a (i)), expressed as percentage of naive model error.

10. Same as in note 9, but with row a (i) replaced by row a (ii).

Naive Model II in column (5). The outcome represents an improvement over Naive Model I, but a relatively small one. The coefficient of relative accuracy in column (5) shows that the record of the anticipatory data is quite good even in terms of this more stringent test.⁹

An additional question frequently raised in assessing forecasts is how well they perform in predicting the direction of change, especially at turning points. By this criterion, obviously, naive model forecasts are very poor. The anticipatory data, on the other hand, appear to have done reasonably well also in this respect. In the course of the eight years covered by the series, which included four turning points, the anticipations for all industries correctly predicted the direction of change in all but one instance, namely the increase of 1950 associated with the Korean War. The manufacturing series missed one additional turn in 1955.

There remains one further refinement of the naive model which is worth considering here. Over the period covered by the investment series there occurred substantial changes in prices and at least part of the fluctuations in the level of investment resulted from such price changes rather than from changes in the physical volume of investment. It is therefore useful to examine the following variant of the naive model test, "Next year's level of investment, valued at current prices, will be the same as the current year's level" (or "the same as the rate of the last quarter of the current year"). Denoting by P an appropriate price index for investment expenditure, the naive model error becomes

$$\frac{I_{t-1} - I_t \frac{P_{t-1}}{P_t}}{I_t \frac{P_{t-1}}{P_t}} = \frac{I_{t-1} \frac{P_t}{P_{t-1}} - I_t}{I_t}.$$

If, instead of extrapolating the current yearly rate it is desired to extrapolate the rate of the last quarter of the year, I_{t-1}^{IV} , the above

formula needs only to be modified by replacing the term $I_{t-1} \left(\frac{P_t}{P_{t-1}} \right)$

9. Even Naive Model II may be slightly biased in favor of the anticipatory data because investment plans are reported toward the middle rather than at the very beginning of the first quarter. A test based on projections of the first quarter rate would be biased, however, in the opposite direction, especially when account is taken of the unavoidable lag in the availability of information about past performance. In any event, rough calculations based on projections of the first quarter rate suggest that the conclusions would not be substantially altered.

by $I_{t-1}^{IV} \left(\frac{P_t}{P'_{t-1}} \right)$, where P'_{t-1} is the price level of the last quarter of the year $t - 1$. The results of such tests, based respectively on the latest year (Model I') and on the last quarter of the latest year (Model II'), are shown in columns (6) and (7).¹

It is apparent that these naive models perform better than the earlier ones, especially when the extrapolation is based on the last quarter. None the less the forecast error remains substantially higher than that of the anticipatory data, two to three times as high when 1950 is omitted. Furthermore, once a price adjustment is introduced in the naive model, it appears proper to inquire whether a similar adjustment could and should also be performed on the anticipatory data. Presumably, what the respondents to the survey are reporting are certain plans for physical investment valued at some level of prices anticipated or assumed.² If this is so, part of the error of the anticipatory data may reflect a failure to anticipate correctly changes in prices rather than a failure to anticipate the correct volume of investment. It would therefore be desirable to try to establish how well the anticipatory data perform in predicting the volume of investments.

To carry out such a test properly it would be necessary to know what assumptions about the future level of prices of investment goods were utilized by the respondents in submitting their investment anticipations. Unfortunately no precise information is available on this point. The question was actually raised in 1956 by the agencies responsible for this survey, and the results of this inquiry have been partly reported in the March 1956 *Survey of Current Business* (p. 20). They show that about one-third of the firms supplying information expected prices to rise, one-third expected prices to remain the same, and one-third did not consider price changes. This information is, unfortunately, of very limited value as it relates to a single year and does not provide any indication of the size of the expected change in prices or the base period to which the change relates.

The authors of the above quoted article seem to consider the base

1. The index P_t for the entire year, and the index P'_t for the final quarter of the year, were constructed from separate deflators for construction and producers' durable goods. The base year for the index is the year 1947. A more detailed description and the actual values of the indexes may be found in the Appendix.

2. This statement controverts one by Messrs. Foss and Natrella in the January 1957 *Survey of Current Business* (p. 18), in which they speculate that some investment plans represent expenditure of the year's depreciation allowance for existing assets.

period to be the time at which the questionnaires are actually filled out, which is the beginning of the year. It seems much more reasonable to suppose, however, that the relevant base period is the planning period, and planning takes place much earlier as it involves making engineering estimates, letting contracts, cash budgeting and arranging for other financing. Hence, if the prices utilized in reporting investment plans were those of the planning period, either because no change in prices was assumed, or because the possibility of changes in prices was disregarded, one might reasonably guess that the price level implicit in the plans is that of the last half of the year. From the information collected in 1956 we know, however, that a sizeable proportion of the respondents takes expected price changes into consideration. It may not be unreasonable to suppose that expected price changes will tend to represent an extrapolation of recent trends. Hence, lacking any precise information, it has been assumed, as a very crude compromise, that the average prices implicit in the reported investment plans are those of the last quarter of the year.³ Needless to say, this is an extremely crude solution. It is used here only for exploratory purposes, pending the availability of better information.

Column (8) of Table I shows the yearly errors of the price-adjusted investment plans. For manufacturing the price adjustment decidedly increases the average forecasting accuracy. The results are less conclusive in the case of all industries. While the average error is reduced somewhat when all years are included, it rises slightly if 1950 is omitted. A comparison of columns (3) and (8) for individual years shows that the price correction almost uniformly reduces the error with one notable exception, namely 1951. In this year the error is increased from 1.8 per cent to 6.7 per cent for all industries, and even for manufacturing it rises from 0 to about 5 per cent. One may well suspect that the poor results for this year reflect the unsatisfactory nature of this price adjustment. In view of the unparalleled increase in prices which occurred throughout most of 1950 it is not unreasonable to suppose that the level of prices implicit in plans for 1951 was above that of the fourth quarter of the year. In this case the price adjustment would overestimate the anticipated level of investment in volume terms, explaining the sizeable overestimate of the price-adjusted anticipatory data. Indeed, if the year 1951 is omitted the average absolute error for all industries falls from 2.6 per cent for the

3. Second half deflators were also tried, but the results were less successful. Another possibility, not attempted, is to use one-half the change between the second half and the following year.

unadjusted anticipatory data to 1.9 per cent for the price-adjusted data. For manufacturing the corresponding decrease is from 2.9 per cent to 1.6 per cent. Finally, it may be noted that the price-adjusted anticipatory data for both manufacturing and all industries predicted the direction of change correctly in every year except 1950.

On the whole, the evidence suggests that if one is interested in forecasting the volume of investment, as is usually the case, one could usefully base such a forecast on the planned *volume* of investment. The planned volume of investment, in turn, could be determined by requesting information from the respondents as to the level of prices implicit in their plans. In the absence of such information, there seems to be some ground for assuming that, on the average, investment anticipations represent the planned volume of investment stated roughly at the level of prices of the last quarter of the year. This assumption will be utilized in the analysis of Section III.

Before closing this section, it is of some interest to inquire whether the anticipatory data exhibit any systematic tendency either to overestimate or to underestimate the level of investment. An answer to this question is provided by the arithmetic mean error of forecast shown in rows *c*(i) and (ii). A positive figure would indicate a tendency for the anticipatory data to overestimate investment, while a negative figure would indicate the opposite tendency. It appears from column (3) that over the years covered by this series, the anticipatory data have systematically tended to underestimate the level of investment. The error is again somewhat greater for manufacturing than for all industries, although if 1950 is omitted it is quite small in both cases. If one could count on a continuation of this tendency, it would be possible to obtain a better forecast from the anticipatory data by adjusting them upward by an appropriate correction factor. However, one should be careful in drawing inferences about systematic biases from the data presented. It must be remembered that, over the period of observation, prices have on the whole tended to rise. The underestimate therefore may merely reflect the failure of the respondents to anticipate this price rise properly. As a matter of fact, if we look at the behavior of the price-adjusted anticipatory data shown in column (8), it is found that, once 1950 is omitted, the mean error is approximately zero for manufacturing and slightly positive for all industries. We may therefore tentatively conclude that the underestimate observed in column (1) reflects primarily the failure to anticipate the rising price trend of the post-war period, rather than any systematic tendency to underestimate the volume of investments.

B. Sales Anticipations

In addition to their investment plans the respondents to the Commerce-SEC survey are also asked to report sales expectations, though not in actual dollars but rather in terms of percentage changes over the previous year. These expectations are analyzed in Table II, which, however, covers only manufacturing inasmuch as for other sectors a continuous series has not been made available by the collecting agencies on the ground that the respondents were too few and hence the estimates unreliable.

The actual level of (monthly average) sales reported in column (1) is taken from the *Survey of Current Business*; the anticipated level of sales in column (2) is computed by applying the anticipated percentage change in sales to the actual level of sales of the previous year. The remaining columns are obtained by an exact repetition of the procedure used in Table I. In order to adjust the sales expectations for price changes one should ideally have information as to the level of prices implicit in the sales forecast. Lacking such information use was again made of the plausible, though not directly verifiable, assumption that, on the average, sales forecasts are stated in terms of prices prevailing in the last quarter of the year. The yearly index of manufactured goods prices is denoted hereafter by π , and the index for the last quarter by π' .⁴

The conclusions suggested by Table II may be summarized as follows:

(a) The average forecast error of sales anticipations is somewhat higher than the error of investment plans, in spite of the smaller variability of the series, as measured by the average error of Naive Model I. Accordingly the coefficient of relative accuracy in terms of this naive model is well below that obtained for investments, though it is still fairly high.

(b) Extrapolation of the fourth quarter rate yields a considerably smaller error, so that, when Naive Model II is used as a yardstick, the coefficient of relative accuracy is reduced further.

(c) Adjustment for price changes diminishes markedly the error of forecast of both naive models but it reduces the error of sales anticipations even more. Accordingly, the coefficient of relative accuracy for the price adjusted anticipations turns out to be fairly impressive, even in terms of Naive Model II.

(d) Sales anticipations appear to have predicted correctly the direction of change in sales in every one of the eight years, which included four turning points. The price-adjusted anticipations were

4. For further details on these indexes, see the Appendix.

TABLE II
ANTICIPATED AND REALIZED SALES AND SELECTED MEASURES OF FORECAST ACCURACY
(MANUFACTURING)

Year	Actual Sales (S_t) (Monthly Average) (\$ Million)	Anticipated Sales ($S_t'(t-1)$) (2)	Forecast Error (%)					
			Raw Data	After Adjustment for Changes in Prices				
			Sales Expectations ¹ (3)	Naive Model I ² (4)	Naive Model II ³ (5)	Naive Model I ⁴ (6)	Naive Model II ⁵ (7)	Sales Expectations ⁶ (8)
1—1947	15,917							
2—1948	17,630	16,713	-5.2	-9.7	-4.6	-2.0	-0.5	-1.1
3—1949	16,416	17,454	6.3	7.4	8.5	5.2	4.2	2.2
4—1950	19,285	16,744	-13.2	-17.4	-18.5	-11.8	-14.5	-8.9
5—1951	22,309	21,406	-4.0	-13.6	-4.5	-4.6	-1.0	-0.6
6—1952	22,850	23,424	2.5	-2.4	-2.5	-4.6	-3.8	1.3
7—1953	24,869	24,450	-1.7	-8.1	-2.1	-7.5	-1.2	-0.7
8—1954	23,396	24,123	3.1	6.3	2.3	6.8	2.2	3.0
9—1955	26,415	24,771	-6.2	-11.4	-11.0	-8.7	-9.2	-4.3
Summary Measures of Forecast Accuracy								
a—Average Absolute Error								
(i) All years			5.3	9.5	6.8	6.4	4.6	2.7
(ii) Excluding 1950			4.1	8.4	5.1	5.6	3.2	1.9
b—Coefficient of Relative Accuracy								
(i) All years				44 ⁷	22 ⁷	58 ⁹	41 ⁹	
(ii) Excluding 1950				51 ⁸	20 ⁸	66 ¹⁰	41 ¹⁰	
c—Arithmetic Mean Error								
(i) All years			-2.3	-6.1	-4.1	-3.4	-3.0	-1.1
(ii) Excluding 1950			-0.7	-4.5	-2.0	-2.2	-1.3	-0.0

1. $\frac{S_t(t-1) - S_t}{S_t} \times 100.$	2. $\frac{S_{t-1} - S_t}{S_t} \times 100.$	3. $\frac{S_{t-1}^{IV} - S_t}{S_t} \times 100.$	4. $\frac{S_{t-1} - S_t}{S_t} \times 100.$	5. $\frac{S_{t-1}^{IV} - S_t}{S_t} \times 100.$	6. $\frac{S_t(t-1) - S_t}{S_t} \times 100.$
---	--	---	--	---	---

$$1. \frac{S_t(t-1) - S_t}{S_t} \times 100. \quad 2. \frac{S_{t-1} - S_t}{S_t} \times 100. \quad 3. \frac{S_{t-1}^{IV} - S_t}{S_t} \times 100. \quad 4. \frac{S_{t-1} - S_t}{S_t} \times 100. \quad 5. \frac{S_{t-1}^{IV} - S_t}{S_t} \times 100. \quad 6. \frac{S_{t(t-1)} - S_t}{S_t}$$

7, 8, 9, 10: see Table I, notes 7, 8, 9, and 10.

equally successful in predicting the direction of change in the volume of sales. At the same time, both the adjusted and the unadjusted anticipations have tended consistently to underestimate the *extent* of change, with the single exception of the year 1952. This tendency confirms the findings of the previous studies of sales forecasts.⁵

(e) Sales anticipations have on the whole tended to underestimate sales, as indicated by a negative arithmetic mean error. But, once more, this result may reflect primarily the failure to anticipate fully the rising trend of prices. Indeed, for the price-adjusted anticipations the average error is rather small, and it vanishes altogether if 1950 is omitted.

In general, then, over the period of observation the forecasting record of sales expectations appears rather good. This conclusion takes on added interest from the fact that previous analyses of sales forecasts collected by other surveys have generally revealed rather poor forecasting performances.⁵ While the present evidence is far too short and limited in coverage to permit firm conclusions, it does suggest that yearly sales anticipations hold some promise as a direct forecasting device.

III. ERRORS OF SALES ANTICIPATIONS AND THE REALIZATION OF INVESTMENT PLANS — AN EXAMPLE OF THE REALIZATION FUNCTION APPROACH

A. *Statement of the Problem*

The present section is concerned with examining whether, and to what extent, the difference between anticipated and actual investments — which in the previous section has been called the error of forecast of investment plans — is affected by the discrepancy between sales expectations and the actual course of sales which is, of course, the error of forecast of sales.

This question is of considerable interest for several reasons. In the first place, both users and analysts of the anticipatory data have suspected that the remarkably good record of these data in forecast-

5. Cf. Modigliani and Sauerlander, *op. cit.*, and R. Ferber, *The Railroad Shippers' Forecasts* (University of Illinois, Bureau of Economic and Business Research, 1953). It should be acknowledged that a recent analysis of Dun and Bradstreet data by M. Hastay, "The Dun and Bradstreet Survey of Businessmen's Expectations," *Proceedings of the Business and Economic Statistics Section* (14th Annual Meeting of the American Statistical Association, Montreal, Sept. 10-13, 1954), pp. 93-123, has attempted to rehabilitate the forecasting record of these data. A forthcoming paper by the authors, however, will cast considerable doubt on Hastay's conclusions.

ing the actual course of capital outlays may have been due to the considerable stability in general business conditions during the period covered by the series. Although the data suggest that, by and large, investment plans have been carried out in the aggregate, one may have doubts as to whether this would continue to be true in the presence of an unanticipated marked business contraction.

It should be possible to throw some light on this question by the kind of analysis suggested above. Although sales were relatively stable over the postwar period, it was shown in Section IIB that there have been repeated, even if moderate, errors in sales expectations, at least for the manufacturing sector, where the information is available. One can therefore examine whether such errors have had a systematic effect on the realization of investment plans. If a stable relation were uncovered between the two variables, this would throw further valuable light on the reliability of investment plans as direct forecasts of investment. Furthermore the stated relation could be of help in improving our ability to forecast general economic conditions, as will be shown in more detail in Section IIIE.

The effect of the disappointment of sales expectations on the realization of investment plans has already been repeatedly and fairly extensively studied at the cross-section level, especially by Bronfenbrenner and Friend and by Eisner.⁶ By and large, these studies have indicated that there exists a positive correlation between the error of sales expectations (actual sales minus expected sales) and the error of investment plans (actual investment minus investment plans). The correlation, however, is not very pronounced and is mostly confined to a certain class of firms which the above authors have referred to as expansionary.⁷ At the time when these tests were carried out, cross-section analysis was the only practicable tool because the anticipatory data covered a very short span of years. With observations now available for eight years it is possible to attempt to carry out a similar analysis at the time-series level even though inferences based on a sample of only eight observations must be regarded as very tentative.

The relation between the actual and planned behavior of the firm on the one hand and the error of expectations on the other has been referred to as "the realization function" in previous studies resulting from the project on "Expectations and Economic Fluctuations."

6. *Op. cit.*

7. Expansionary firms may be roughly characterized as firms whose investments were largely for purposes of expansion, the classification between expansionary and nonexpansionary firms being, of course, based on criteria other than the responsiveness to the error of forecast.

The nature of this function has been examined at some length elsewhere⁸ and therefore will be discussed only briefly here. It has been shown that there are good a priori reasons for expecting that the discrepancies between actual and planned behavior should largely be accounted for by the discrepancy between the behavior of the environment as anticipated at the time the plan was made and the actual behavior of the environment over the period to which the plan refers. By "environment behavior" is meant, in this connection, the behavior of all variables which are outside the direct control of the firm, as, for instance, the behavior of demand for the firm's product, prices to be paid for factors of production, the cost of securing capital, etc.

The asserted relation between the realization of plans and the error of expectations is supported by the consideration that a plan does not represent a decision by the firm as to what it will do at various future points of time. Decisions, in this sense, are made only at the time when they are to be carried out. A plan should rather be regarded as the best guess that the firm can now make as to what future course it will adopt *if* the expectations held at the time of planning have not changed significantly. Such plans are made not for the purpose of deciding *future* actions, but rather for the purpose of deciding the best *current* course of action. Thus a plan should be regarded as a conditional statement of what the firm will do, *if* the expectations are not revised in subsequent periods. The actual course of action, however, will eventually depend on the behavior of the environment over the planning period. If the behavior of the environment departs from expectations, the actual behavior of the firm will also tend to depart from plans. Thus, by and large, the discrepancy between actual and planned behavior should be a function of the discrepancy between the actual behavior of the environment and the anticipated behavior — the discrepancy tending to be smaller in absolute value, the smaller the error of expectations.

In principle, the realization function should include all variables that may exert a significant influence on the actual behavior of the firm and which are capable of significant unanticipated variation over the interval covered by the plan. In general, however, it need not include variables which may be expected to exert an influence on the behavior of the firm if these are not capable of variation within the

8. F. Modigliani and K. J. Cohen, "The Significance and Uses of *Ex Ante* Data — A Summary View," *Proceedings of the Conference on Expectations, Uncertainty, and Business Behavior*, Social Science Research Council, Oct. 27-29, 1955, forthcoming, and *The Role of Anticipations and Plans in the Economy of the Firm, and Their Use in Economic Analysis and Forecasting*, Bureau of Economic and Business Research, University of Illinois, Part II.

interval. That is, it need not include initial conditions. This is because the initial conditions are fully reflected in the plan itself, as well as in the anticipations underlying the plan.

In the present case the only information available on relevant expectations refers to sales. A priori considerations, as well as the results of the cross-section studies referred to above, suggest that these are not the only relevant variables. It would seem, nevertheless, that sales are a sufficiently important variable to justify formulating and fitting a realization function for investment involving this anticipatory variable alone.

There remains the problem of determining the exact form of the relation to be tested, which is taken up in the next section.

B. Formulation of the Model

In view of the limited information available and the shortness of the series, the model about to be presented has been kept to a maximum of simplicity, leading also to simple empirical tests.

The point of departure is a well-known variant of the acceleration hypothesis.⁹ Specifically, the level of investment planned at the end of year $t - 1$ for the year t is taken to be a fraction of the difference between the desired stock of capital, as seen at the planning date, and the existing stock adjusted for wear and tear in the course of the year. This hypothesis can be stated formally as follows:

$$(1) \quad I_t(t - 1) = \gamma [K_t^d(t - 1) - (1 - \delta)K_{t-1}]$$

The three variables appearing in this equation, $I_t(t - 1)$, $K_t^d(t - 1)$ and K_{t-1} represent, respectively, planned investments for the year t , the desirable stock of capital and the actual stock of capital, as of the planning date, all regarded as measured in "real" terms, i.e., at prices of some chosen base year. The parameter δ denotes the fraction of the initial stock expected to be worn out in the course of the year t and γ is a coefficient measuring the speed with which firms, on the average, endeavor to adjust the stock of capital. Both δ and γ , as well as the remaining coefficients about to be introduced, will be regarded as approximately constant over the period of time relevant for the analysis.

In line with the notion of the acceleration principle, the optimum stock of capital as seen at the planning date can be expressed as

9. For a more detailed discussion of the basic model underlying equations (1) and (2) below, see F. Modigliani, "Comment on 'Capacity, Capacity Utilization, and the Acceleration Principle' by B. C. Hickman" in *Problems of Capital Formation* (*Studies in Income and Wealth*, XIX, National Bureau of Economic Research, Princeton, 1957).

$$(2) \quad K_t^d(t-1) = \alpha S(t-1)$$

where $S(t-1)$ denotes the rate of sales which the firm should be geared to meet, as anticipated at the planning point, and α is the so-called capital coefficient (or acceleration coefficient).¹ Substituting from (2) into (1) yields

$$(3) \quad I_t(t-1) = \gamma[\alpha S(t-1) - (1-\delta)K_{t-1}].$$

By analogous reasoning, the optimum stock of capital, as seen in the light of conditions actually prevailing in the year t — the *ex post* optimum as we may call it — will be

$$K_t^d = \alpha S(t)$$

where $S(t)$ represents the revised anticipated rate of sales. Therefore the “*ex post* desirable” level of investment, say I_t^* , can be expressed as

$$(4) \quad I_t^* = \gamma[\alpha S(t) - (1-\delta)K_{t-1}].$$

The actual level of investment in the year t can then be expected to represent a weighted average of the initially planned investment and the *ex post* desirable investment, i.e.,

$$(5) \quad I_t = \beta I_t^* + (1-\beta)I_t(t-1) = I_t(t-1) + \beta[I_t^* - I_t(t-1)]$$

where β may be regarded as a measure of the speed with which initial plans for the given year are modified within the year itself in the light of information accruing in the course of the year. Now, making use of (3) and (4), the quantity in brackets in (5) can be reduced to

$$I_t^* - I_t(t-1) = \gamma[\alpha S(t) - (1-\delta)K_{t-1}] - \gamma[\alpha S(t-1) - (1-\delta)K_{t-1}] = \gamma\alpha[S(t) - S(t-1)].$$

Substituting this result into (5) yields

$$(6) \quad I_t = I_t(t-1) + \beta\gamma\alpha[S(t) - S(t-1)].$$

No direct information is available on the variables $S(t)$ and $S(t-1)$ appearing in the above equation. We do, however, have information on the level of sales anticipated for the near future, $S_t(t-1)$ and we may conjecture that the change in the relevant long-run sales expectations will largely depend on the extent to which

1. The coefficient α is, in general, not determined exclusively by technical considerations. It may include, for instance, appropriate allowances for a *normal* margin of spare capacity related to the short-run variability and the long-run growth trend of sales. See on this point the reference cited in the preceding note and also H. B. Chenery, “Overcapacity and the Acceleration Principle,” *Econometrica*, Vol. 20 (Jan. 1952), and A. Kisselgoff and F. Modigliani, “Private Investments in the Electric Power Industry and the Acceleration Principle,” *Review of Economics and Statistics*, XXXIX (Nov. 1957).

these shorter-run anticipations have been confirmed, or have failed to be confirmed, by the actual course of sales. We propose, therefore, to approximate the quantity $S(t) - S(t - 1)$ in terms of observables as follows:

$$(7) \quad S(t) - S(t - 1) = \mu[S_t - S_t(t - 1)],$$

where the proportionality factor μ is akin to the Hicksian elasticity of expectations. From (7) and (6) we now derive

$$(8) \quad I_t = I_t(t - 1) + b[S_t - S_t(t - 1)], \quad b = \mu\beta\gamma\alpha,$$

which represents the desired realization function, expressing investment as a function of investment plans and the sales forecast error. It will be noted that in this equation the initial stock of capital, K_{t-1} , no longer appears. This variable, in spite of its obvious influence on investment (which is recognized explicitly in equations (3) and (4)), is an initial condition, and therefore does not appear explicitly in the realization function. Its influence on investment is exerted entirely through planned investment, which depends, of course, on the initial stock of capital, as well as on other initial conditions.²

Equation (8) is not yet in a form entirely suited for empirical testing because all the variables are measured in prices of some base year, whereas the survey data to be utilized in the tests are measured in current prices. However, equation (8) can readily be expressed in terms of current prices. Denote respectively by P_t and π_t the average value of the price index of investment goods and of finished products in the year t relative to the base year. Similarly denote by P'_{t-1} and π'_{t-1} the average level of prices relative to the base year implicit in

2. For the sake of simplicity the reasoning leading up to the realization function, equation (8), has glossed over a number of technical details necessary for a more rigorous and general formulation. The desired stock of capital is presumably affected by initial conditions other than the initial stock of capital, e.g., by balance sheet variables, and by anticipatory variables other than sales expectations, e.g., the availability and cost of funds. If these variables appeared in equations (2) and (4) in additive form, the initial conditions would again drop out in the realization function, although additional terms expressing disappointment of other expectations would appear. (Some of these disappointment terms may represent the difference between initially held expectations and expectations held at a later point of time that relate to the more distant future.) Initial conditions not entering in linear form, by contrast, may appear as parameters in the realization function. Such refinements do not appear warranted for the present analysis in view of the nature of the data available at the present time.

Finally, it must be recognized that the various hypotheses stated in the text are of a stochastic nature, and accordingly each of the corresponding equations should have contained a random term. This is especially true since, in the formulation given, many relevant variables have been omitted; their effect must be regarded as subsumed in the random term. These random terms have been omitted in the text for brevity.

the investment plans and in the sales forecast for the year t . Then equation (8) can be rewritten as

$$(9) \quad \frac{I_t}{P_t} = \frac{I_t(t-1)}{P'_{t-1}} + b \left[\frac{S_t}{\pi_t} - \frac{S_t(t-1)}{\pi'_{t-1}} \right]$$

where I_t and S_t again represent investment and sales in current prices and similarly for the corresponding anticipations. If we now multiply both sides of this equation by P_t , and for brevity let

$$X_t = I_t(t-1) \frac{P_t}{P'_{t-1}}, \quad E_t = \left[S_t - S_t(t-1) \frac{\pi_t}{\pi'_{t-1}} \right] \frac{P_t}{\pi_t},$$

hypothesis (9) can be conveniently rewritten as

$$(10) \quad I_t = X_t + bE_t,$$

which can be generalized to

$$(10') \quad I_t = aX_t + bE_t + c$$

to allow for the possibility of systematic biases in investment plans, as discussed below. Hypothesis (10') can be tested directly from available Commerce-SEC data, if one is willing to approximate P'_{t-1} and π'_{t-1} by the corresponding price indexes for the last quarter of the previous year, as suggested in Section II.

In addition to the hypothesis just developed an alternative formulation will also be tested in the next section. The reason for formulating a variant of the hypothesis, which is given in equation (12') below, is that estimates of the parameters of equation (10') obtained by the method of least squares may be subject to some bias because of the sizable range of the dependent variable.³ This danger can be reduced by dividing all terms by a suitable scale factor. In previous analyses of data of this type at the cross-section level the value of fixed assets has been commonly used as the divisor.⁴ Unfortunately this method cannot be utilized here because reliable information on aggregate fixed assets is not readily available. Hence the following alternative procedure has been adopted. One can first transpose the planned investment term of equation (9), obtaining

$$(11) \quad \frac{I_t}{P_t} - \frac{I_t(t-1)}{P'_{t-1}} = b \left[\frac{S_t}{\pi_t} - \frac{S_t(t-1)}{\pi'_{t-1}} \right].$$

3. It cannot be assumed that the data have the property of homoscedasticity, or constant variance, throughout the range. It seems far more likely that the variance is proportional to the expected value of the dependent variable.

4. Friend and Bronfenbrenner, *op. cit.*, and Eisner, *op. cit.*

Now it is possible to divide each side by either variable found there, to wit, I_t/P_t or $I_t(t-1)/P'_{t-1}$ on the left, and S_t/π_t or $S_t(t-1)/\pi'_{t-1}$ on the right. In particular, selecting the anticipatory variables for the divisors, one obtains⁵

$$(12) \quad \frac{(I_t/P_t) - (I_t(t-1)/P'_{t-1})}{I_t(t-1)/P'_{t-1}} = b' \left[\frac{(S_t/\pi_t) - (S_t(t-1)/\pi'_{t-1})}{S_t(t-1)/\pi'_{t-1}} \right].$$

In this equation the left-hand side expression (multiplied by 100) represents the percentage error of investment plans, denoted hereafter for brevity by I'_t , while the expression in brackets on the right hand (multiplied by 100) is the corresponding percentage error of sales expectations, denoted hereafter by E'_t . Hypothesis (12) can again be broadened to

$$(12') \quad I'_t = b'E'_t + c'$$

and a positive (negative) value of the constant term can readily be interpreted as indicating a systematic downward (upward) percentage bias of investment plans.

While equation (12') has certain statistical advantages over (10'), it also has certain drawbacks in that there is no clear analytical justification for dividing the two sides of (11) by two different quantities. None the less, hypothesis (12') is clearly a sensible one and has the further advantage, as compared with that of equation (9), that it involves only two variables whose relation can be exhibited graphically on a scatter diagram.

C. Empirical Results

Proceeding to the empirical test of the hypotheses, the following results are obtained for equation (10'):

$$(13) \quad I_t = 0.91X_t + 0.031E_t + 0.98, \quad R = 0.989^a$$

$$(\pm 0.05) \quad (\pm 0.011)$$

(all variables measured in billions of dollars).⁷

5. The choice of the anticipatory variable for the divisor is dictated by considerations of convenience for forecasting purposes. While the left side of (12) does not directly yield a forecast of investment, as is true for equation (9), a forecast can be obtained very simply by first multiplying both sides of equation (12) by $I_t(t-1)/P'_{t-1}$ which is known in advance, and then adding this same quantity to the result. This manipulation immediately yields the forecast of the volume of investment, I_t/P_t . This volume forecast can then be expressed in terms of any desired price level by multiplying by the corresponding price index. In particular, one can obtain a forecast of the value of investment in the following year by multiplying by the forecasted level of prices.

6. Adjusted for degrees of freedom.

7. It should be noted that, in the computation of the variable E_t , S_t and $S_t(t-1)$ were measured in terms of yearly sales rather than average monthly

In terms of standard statistical criteria these results are rather favorable to the hypothesis presented. Focusing on the critical variable for this test, namely E , it is found that its regression coefficient "passes" the customary test of significance at the 5 per cent level. In other words, the error of sales expectations appears to exert a significant effect on the relation between investment plans and actual investment. Another view of these results can be gained from Table III which presents actual investments together with the value calculated from (13) and the corresponding hypothetical percentage

TABLE III

	I_t^c	I_t	$\frac{I_t^c - I_t}{I_t} \times 100$
	(1)	(2)	(3)
1948	9,176	9,134	0.5
1949	7,260	7,149	1.6
1950	7,401	7,491	-1.2
1951	11,351	10,852	4.6
1952	11,605	11,632	-0.2
1953	11,821	11,908	-0.7
1954	10,694	11,038	-3.1
1955	11,338	11,439	-0.9
Average Absolute Error			1.6

error of forecast. It appears that (13) provides a remarkably accurate "explanation" of investment for most years, including 1950. The most noticeable exception is the year 1951 when, as noted earlier, the error may have resulted from an overcorrection for price changes; presumably more direct information on prices implicit in the anticipations would reduce such errors. A comparison of column (3) with column (8) of Table IA indicates that, with the help of (13), one can account for nearly 60 per cent of the average discrepancy of 3.8 per cent between actual investments and the price-adjusted investment plans.

A further bit of evidence for the usefulness of this approach is provided by the results of extrapolating this equation to the year 1956, which was not included in estimating the regression coefficients.

sales, as in Section IIB. Also, in estimating (13) the year 1950 was not excluded for its omission does not seem appropriate here, even though it was appropriate in the analysis of Section II. The point is that, in the present section, the error in sales anticipations is taken into account and it is not apparent that the effect of a forecast error resulting from an unanticipated military emergency is necessarily qualitatively different from the effect of errors arising from other unanticipated events.

In this year manufacturers' investment plans, adjusted for price changes, amounted to \$15.7 billion, an overestimate of 4.9 per cent over actual outlays of the year, which came to \$15.0 billion. However, sales expectations overestimated actual sales by 3.9 per cent (after adjustment for price changes), or by \$12.9 billion out of \$321 billion. Utilizing these values in equation (13) yields a computed investment of \$14.8 billion, which represents an error of only 1.2 per cent.

The numerical estimates of the coefficients of (13) also deserve some comments. The coefficient of X_t is seen to fall short of unity, and fairly significantly so, when account is taken of its standard error. This finding, per se, would imply that investment plans tend systematically to overestimate investment by approximately 9 per cent, even when there is no error of sales expectations. This apparent tendency is, however, largely offset by the constant term which implies an underestimation also roughly on the order of 10 per cent. This combination of a relatively low coefficient of X_t offset by a relatively high constant term indicates that plans do not, on the average, tend to over- or underestimate the *level* of investment — as was already suggested by the analysis of Section II — but that they *do* tend to overestimate systematically the extent of year to year *change*.⁸ This tendency is diametrically opposite to that noted for sales anticipations.

Proceeding to the test of the alternative hypothesis (12'), one obtains the following results

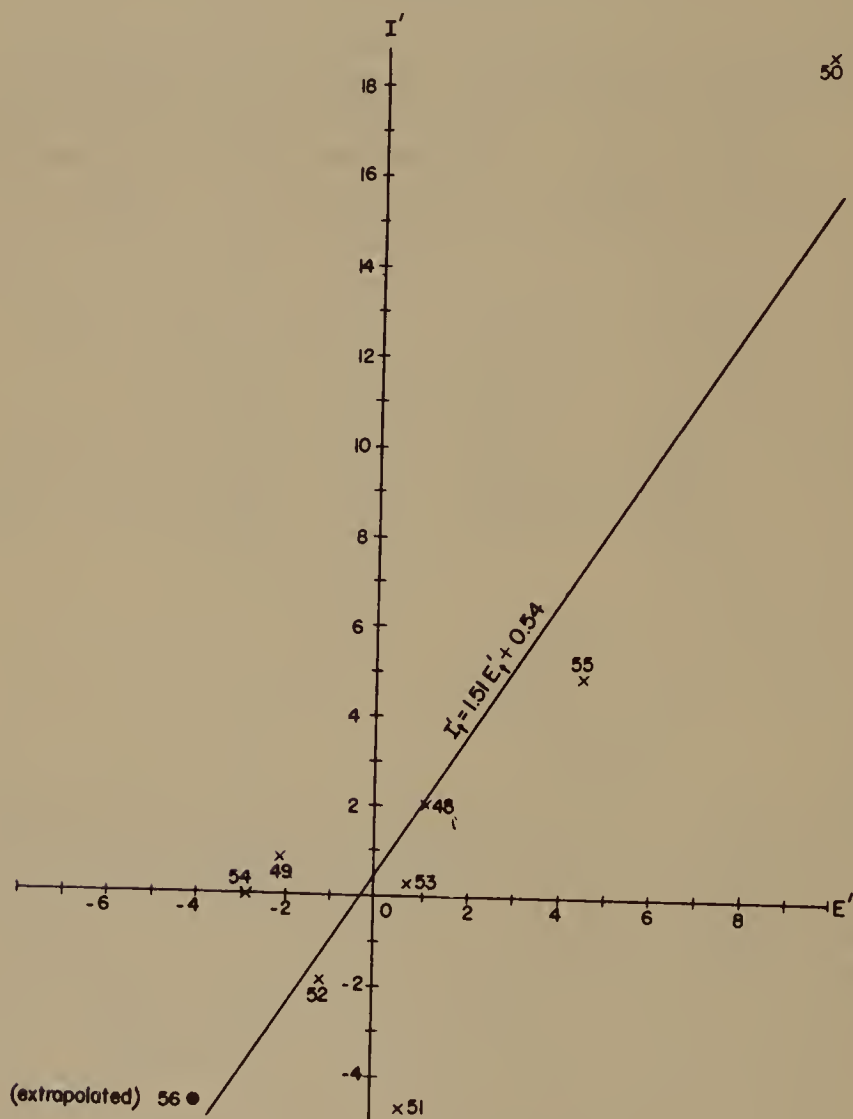
$$(14) \quad \begin{aligned} I'_t &= 1.51E'_t + 0.54 & r &= 0.85 \\ &(\pm 0.34) \end{aligned}$$

which can again be regarded as quite satisfactory, the correlation coefficient being significant, by standard criteria, at the 1 per cent level. The relation between I' and E' is also exhibited in the scatter diagram of Figure I, together with the regression line. The positive estimate for the constant term indicates that plans have a systematic downward bias, but the indicated bias is negligibly small — about one half of 1 per cent — which is in line with previous results. Also, according to the regression coefficient, a 1 per cent error in sales

8. This tendency can be detected even by a direct comparison of actual with anticipated investment corrected for price changes, and can be inferred from Table IA by comparing the figures of column (8) with those of column (6). It will be seen that in the five years in which investment rose by more than 3 per cent the anticipated change was even larger (or at least as large). The same phenomenon appears to have been repeated in 1956. For the unadjusted data of column (2) this phenomenon is not readily apparent because it is hidden by the prevailing underestimation of price changes.

expectations tends to generate an excess of investment over plans by roughly 1.5 per cent. As the scatter shows, these results are heavily influenced by the extreme observation of 1950. It does not seem, however, that they are entirely due to this extreme observation. Indeed, if one is willing to discount the observation for the

FIGURE I



$$I'_t = \frac{(I_t/P_t) - (I_{t-1}/P'_{t-1})}{I_{t-1}/P'_{t-1}} \times 100$$

$$E'_t = \frac{(S_t/\pi_t) - (S_{t-1}/\pi'_{t-1})}{S_{t-1}/\pi'_{t-1}} \times 100$$

year 1951 for the reasons repeatedly mentioned, the remaining points are seen to fall roughly into a line with the year 1950. Furthermore, the observation for the year 1956, again not utilized in the regression, falls almost exactly on the regression line.⁹

D. Some Analytical Implications

The empirical results of the previous section provide some interesting though tentative insights into the influence of current business conditions on investment. In the first place, these results indicate that realization of plans, and hence the actual level of investments during a given year, is indeed influenced by the course of economic activity in the same year, at least to the extent that this course affects the behavior of current sales. (Current developments may, of course, influence the realization of plans through other links, though this possibility cannot be analyzed at present, lacking anticipatory data other than those on sales.) Specifically, equation (9) is seen to imply that the real volume of investment in a given year, I_t/P_t , is a linear function of the volume of sales in the same year, S_t/π_t , since the remaining two variables, investment plans and sales expectations, are given as of the beginning of the year, and therefore are not subject to change during the course of the year. By comparing equation (9) with equations (10) and (13), it is seen that a quantitative estimate of the effects of current sales on investment is provided precisely by the coefficient of the variable E in (13), namely, 0.031. Since both investment and sales are measured in billions of dollars at 1947 prices, this coefficient implies that a change of \$1 in current yearly sales tends to change current investment by about 3 cents in the same direction.¹

9. Similar tests were also carried out for the data without adjustment for price changes. The equation corresponding to equation (13) in the text is

$$(13') \quad I_t = 0.91 I_t(t-1) + 0.026 (S_t - S_t(t-1)) + 1.02. \quad R = 0.995$$

(± 0.04) (± 0.005)

The multiple correlation coefficient, as well as the standard errors of the regression coefficients, would seem to indicate that the unadjusted data yield "better" results. However, for the reasons cited in the text, there are strong a priori reasons for regarding the price-adjusted model as more meaningful, and hence more reliable in the long run. Actually, for 1956 the predicted value of investment computed from equation (13') is \$14.64 billion, which represents an error of 2.1 per cent, considerably poorer than the error of 1.2 per cent obtained from the extrapolation of equation (13). The counterpart to equation (14) which omits the price adjustments is

$$(14') \quad I_t'' = 1.03 E_t'' + 1.79. \quad r = 0.88$$

(± 0.21)

Comments similar to the above apply here also.

1. This statement applies strictly only for the volume of investment and sales measured in base year prices, i.e., in the present instance, 1947 prices.

This estimate can be compared with the results obtained for the alternative formulation of equation (14). According to this equation, a 1 per cent error in sales forecasts tends to alter investment by 1.5 per cent of the planned figure. To convert this relation between per cent errors to a relation between actual investment and actual sales, one needs to multiply this 1.5 per cent by the ratio of investment to sales.² While this ratio has, of course, varied from year to year, it is of the order of $1/25$, indicating that a change of \$1 in current sales has tended to affect investment by $1.5/25 = 0.06$, or 6 cents per dollar. This estimate is somewhat larger than that provided by equation (13), but it is not inconsistent with that estimate when account is taken of the effect on both estimates of sampling fluctuations, as shown by the standard errors.

These results suggest that although current sales have a systematic effect on current investment, this effect is quantitatively rather small, especially as compared with what one might expect from the acceleration principle in its original strict formulation. According to this formulation, the stock of capital adjusts instantaneously to the *rate* of sales so that investment is proportional to the *rate of change* of sales. The factor of proportionality is the capital coefficient, the α of equation (2), which represents the ratio of the optimum stock of capital to yearly sales. If this model were correct, an increase in the yearly rate of sales above the expected rate should lead to an increase in investment over plans equal to the unexpected increase in the rate of sales times the capital coefficient. In other words, the coefficient of the variable E in equation (13) would be equal to the capital coefficient α .

There is very little statistical information on the size of the capital coefficient for manufacturing industries. All the information that is available, however, indicates that it is a great deal larger than the coefficient of 0.03 of equation (13) or even the larger estimate of 0.06 derived from equation (14). For example, the actual ratio of the book value of fixed assets to the value of annual sales in recent

However, it can be seen from equation (10) and the definition of E_t , that this relation is approximately valid also in terms of current prices inasmuch as the ratio P_t/π_t is generally close to unity, prices of investment goods and finished products generally tending to move closely together.

2. The regression coefficient of equation (14), say b' , measures essentially the elasticity of investment with respect to sales, i.e., $\frac{\Delta I}{I} / \frac{\Delta S}{S} = \frac{\Delta I}{\Delta S} \times \frac{S}{I}$. Therefore the desired estimate of $\frac{\Delta I}{\Delta S}$ is given by $b' \frac{I}{S}$.

years has been of the order of 45 cents per dollar of yearly sales.³ Furthermore, this figure undoubtedly underestimates the capital coefficient substantially in that in a period of rising prices, book value (the only measure available) is considerably below replacement cost, which is the relevant measure.⁴

The finding that the effect of a change in current sales on investment is very much smaller than any reasonable estimate of the capital coefficient is, of course, fully consistent with our model according to which the coefficient of the variable E is not simply α but represents instead the product of α and three other coefficients. Each of these coefficients may be expected to be well below unity: γ , because the adjustment in the stock of capital will be planned to occur but gradually over time; μ , because current sales may be expected to have but a limited influence on the longer-run sales outlook which controls the stock of capital that is eventually to be secured; β , because of the difficulty of revising plans in the short run, even when expectations have been revised in the light of current events.

Unfortunately the analysis does not per se provide a direct numerical estimate of the individual coefficients. It can be used, however, to shed at least some light on the possible value of the critical product $\mu\beta$ which measures the effect of current sales experience on the current rate of investment. Clearly this product must be larger than zero, for, if it were zero — as would happen if plans were absolutely rigid one year ahead or the behavior of current sales had no effect at all on longer-run anticipations — then the coefficient of E would itself be zero. But to draw more specific inferences would require quantitative information not only about α but also about γ . Unfortunately, very little is known at present about γ . Some very fragmentary evidence relating to specific manufacturing industries suggests that it might be on the order of one-third to one-fifth.⁵ If then the value of α is estimated at between .5 and 1, which is probably still a conservative estimate, it would appear that the value of $\mu\beta$ is not likely to be larger than .5, and it is probably considerably lower, possibly on the order of .1 or even less.

3. Based on *Quarterly Financial Reports, United States Manufacturing Corporations*, Federal Trade Commission and Securities and Exchange Commission.

4. Some information on capital coefficients is being provided by the so-called interindustry studies or input-output analyses. This information cannot be utilized for present purposes because of obvious aggregation problems. However, Harold J. Barnett, in his comment to Hickman, *op. cit.*, p. 465, suggests that the capital coefficient might be closer to 2.

5. See F. Modigliani, "Comment on 'Capacity, Capacity Utilization, and the Acceleration Principle,'" *op. cit.*

In summary, the evidence available at present suggests that, on the average, both investment plans and long-run expectations tend to be fairly rigid in the short run, with the result that the effect of current sales on current investment, though not entirely absent, is rather minor. Furthermore, the results of Table III suggest that the effect of current events other than sales is also not likely to be very pronounced.⁶ These conclusions must, of course, be regarded as extremely tentative. In the first place, they are based on very few observations and on very uncertain estimates of α and γ . Second, and more important, over the period of observation there is but a single instance in which sales were significantly above expectations — namely the year 1950 which can hardly be regarded as representative — and no single instance in which sales fell significantly short of expectations. It is quite conceivable that in the presence of a substantial disappointment of sales expectations there would be considerably more readiness to adjust plans downward than is suggested by the observations presently available.

On the other hand, if the apparent short-run unresponsiveness of investment is confirmed by further analysis, this finding would have rather far-reaching analytical and policy implications. From an analytical point of view, for instance, it would imply that the level of investment over intervals on the order of a year or less can be explained fairly closely in terms of lagged variables alone. From the policy point of view it would strengthen the case of those who doubt the short-run effectiveness of monetary and credit policies in controlling the rate of investment in fixed capital.⁷

E. Some Forecasting Implications

A few remarks are still needed concerning possible forecasting applications of realization functions of the type represented by equa-

6. It may be suggested that this unresponsiveness of investment to current events merely reflects the fact that commitments involved in investment in plant and equipment must be made with long lead times, and that plans simply report these commitments. It would follow, if this were the case, that all one would need for accurate forecasts of investment is information on commitments. While this point could never be refuted by a time-series analysis, it is a type of problem on which cross-section evidence can shed much light. The analysis of this evidence (see references in note 3, p. 24) would seem to cast considerable doubt on this proposition. In any event, the recently initiated survey of capital appropriations by the National Industrial Conference Board should soon provide further evidence on this point.

7. Note that even if this conclusion is valid, it does not imply that monetary policy is altogether ineffective in influencing the level of investments. It merely indicates that the effectiveness of such policies is not very significant in the short run, i.e., for spans on the order of a year. However, monetary and credit policy may still have a significant lagged effect by affecting investments for later periods.

tions (13) and (14). As indicated earlier these equations do not directly yield a forecast of investment in the coming year. While two of the variables appearing in the equations, namely $I_t(t-1)$ and $S_t(t-1)$, are known at the beginning of the year, to make a forecast one still needs to know the level of sales for the year, S_t .⁸ Under these circumstances it would appear as though these equations are of little help in forecasting, for they require a prior forecast of S_t , while generally the main purpose of forecasting I_t is precisely that of arriving at a forecast of the level of activity, of which S_t is itself a good measure.

In spite of this circularity, realization functions of the type of equations (13) and (14) can be effectively exploited for forecasting purposes by *imbedding them in broader models which explicitly or implicitly take into account the interdependent nature of economic variables*. Such models may either take the form of a closed econometric system of simultaneous equations of the type made familiar by the work of Tinbergen, Klein and others,⁹ or they may take a less formal structure as those underlying many variants of the extensively used Gross National Product approach.¹

The way in which a relation of the type of equation (13) can be utilized in either of these approaches can be sketched out by taking a very simple illustration. One may begin with equation (13) (although what follows is just as applicable if one used equation (14) as an alternative) by assuming tentatively that the level of sales will be exactly the level anticipated. Since E is then equal to zero, the tentative forecast of investment can be computed, utilizing only known information on plans. From this computed investment, and certain other relations such as the relation between consumption and income, i.e., a consumption function, one can then compute the level of income (and of consumption) which would correspond to the given level of investment. From the known level of income one can project the

8. For the purpose of the present discussion the assumption is made that prices remain at the level of the previous year, or, equivalently, that consideration is limited to forecasting the volume of investment. In order to forecast the value of investment it would be necessary in addition to be able to forecast changes in the price level. In what follows, therefore, the symbol I_t is used to denote the volume of investment, i.e., in notation used previously, I_t/P_t , and similarly for the remaining symbols.

9. Cf. J. Tinbergen, *Statistical Testing of Business Cycle Theories*, Volume 2, "Business Cycles in the United States of America, 1919-1932," League of Nations, Economic Intelligence Service (Geneva, 1939); L. Klein, *Economic Fluctuations in the United States*, Cowles Commission Monograph No. 11 (New York, 1950).

1. See, for instance, E. C. Bratt, *Business Cycles and Forecasting* (Homewood, Ill., 1953), especially chaps. 17 and 18, and the references cited therein.

level of sales which would be generated by this level of income.² One can then compare this computed level with sales anticipations. If the value of computed sales agrees with the forecast of sales, the value of E will be zero, indicating that the tentative forecast of investment is consistent with the sales implications of this investment plan and that, therefore, the initial forecast need not be adjusted further. The final forecast has then been reached. If, on the other hand, E is different from zero, a second estimate of investment can be made by utilizing equation (13) again. From this estimate a new estimate of income and of sales, and a new value for E may be obtained, which can again be compared with its initial value. This procedure can be repeated until one finds a value of E such that the level of investment corresponding to it generates a level of sales equal to that implicit in the initial value of E .

This iterative procedure is, of course, not necessarily the computational method one would apply. The actual process utilized would depend on the specific nature of the over-all model, which in practice would be considerably more complex than that described above. In the special case in which all the relevant relations can be expressed in terms of equations, the iterative process can be replaced by simultaneous solution of the system, equation (13) (or (14) appearing as one of the equations of the system.

At the present time the possibility of utilizing the approach just illustrated is limited by the nature of the published information. The availability of data on other expectations and on other sectors of the economy, should make it possible to refine the analysis and extend it to such other sectors, increasing the range of practical applicability of the realization function approach.

In summary, in the short span of years for which the information is available, yearly investment plans and sales expectations have had a very favorable record as direct forecasts, especially after making allowances for the effect of changes in the general level of prices. This result is especially significant in the case of investment plans since alternative approaches have not proved very successful in forecasting investment in fixed capital, which is, in turn, a key variable in short-run forecasting of economic activity. The usefulness of these anticipatory data may be further enhanced by the simultaneous use of expectations and plans in the realization function. This approach enables us to exploit anticipatory data without losing sight of the

2. The level of sales will most likely be some multiple of income, since sales is a gross concept, involving a great deal of double counting.

interdependent nature of economic variables and holds some promise as a useful addition to the tool kit of the short-run forecaster.

APPENDIX

The price indexes used to deflate plant and equipment expenditures were derived from the *Survey of Current Business*, and from the *National Income Supplement to the Survey of Current Business*. The annual index for all but the last year represents a weighted average, in the ratio of two to one, of the "Producers' Durable Equipment" Deflator and "Other New Construction" Deflator of Gross National Product. This ratio represents the approximate relation between the value of equipment purchases and plant construction during this period.

The fourth quarter index for this series was constructed by interpolating the annual index with the aid of the wholesale price index of "Total Machinery and Motive Products," and the E. H. Boeckh Associates index of construction costs for commercial and factory buildings — brick and steel — from the *Survey of Current Business*. A similar procedure was used to estimate the annual index for 1956.

The deflator for sales of the manufacturing sector which was used is the wholesale index of "Commodities Other than Farm Products and Foods" from the *Survey of Current Business*.

The values of the indexes are given in the table below.

TABLE IV

PRICE DEFLATORS FOR INVESTMENT AND SALES

	Investment Deflators 1947 = 100 Annual (P_t)	Sales Deflators 1947-49 = 100 Annual (π_t)
1947	100.0	95.3
1948	109.0	103.4
1949	112.9	101.3
1950	115.2	105.0
1951	125.1	115.9
1952	126.6	113.2
1953	129.3	114.0
1954	129.9	114.5
1955	132.1	117.0
1956	140.9	122.2

	4th Quarter (P'_t)	4th Quarter (π'_t)
1947	103.4	99.1
1948	113.7	105.4
1949	112.1	100.1
1950	119.4	111.9
1951	125.7	114.6
1952	127.1	112.9
1953	130.2	114.6
1954	129.7	114.7
1955	135.2	119.4

F. MODIGLIANI.

HARVARD UNIVERSITY.

(On leave from Carnegie Institute of Technology)

H. M. WEINGARTNER.

UNIVERSITY OF CHICAGO

ECONOMIC PLANNING RECONSIDERED

By JOHN E. ELLIOTT

Definitional problems, 55. — The structure of economic planning, 61. — The process of economic planning, 67. — Summary and conclusions, 75.

DEFINITIONAL PROBLEMS

Introduction. It hardly needs to be said that “planning” or “economic planning” means many things to many people, and has been used, in professional and popular discussion alike, in widely divergent ways, and in a highly emotional manner. Because of the arbitrary and personal nature of such discussions, and because of the well-known difficulty of achieving anything even remotely approaching resolution of such issues, there is little to be gained from an intensive review, comparison, and assessment of myriads of alternative possible definitions.

Further, a definitional approach to economic planning, in addition to this practical difficulty, involves or may involve serious methodological limitations, including among others these two:¹ First, the attempt at the statement of the essence of a problem or issue in a formalized definition is opposed to the nominalistic and empirical temper of modern scientific research. Clarity of concepts is, of course, a prerequisite to systematic endeavor in any field of study. But as Marshall and J. M. Keynes among others have reminded us, these concepts are not themselves truth or reality; they are rather tools of analysis, ways of looking at problems, and so on. A definitional approach to economic planning can easily descend into scholasticism and nonempirical verbiage.

Second, the attempt to define economic planning assumes that this complex and many-sided phenomenon is specific and concrete enough to be defined. It is here submitted that economic planning is a complex clustering of problems to be “explored,” not “defined.”

Still, the renewed and lively interest in the methodology of economic planning, policy, and welfare in recent years² would seem to point toward the possibility and desirability of restating the basic elements of this concept. In short, “economic planning,” despite the abuses to which it has been subjected, is, or at least could be, a useful

1. On this, and other, issues, concerned with the methodology of planning, the writer is indebted to discussions with Professor Arthur Maass of Harvard University.

2. See, e.g., Arrow, *Social Choice and Individual Values*; Baumol, *Welfare*

phrase and concept. From a definitional point of view, the problems are: (1) to delineate the nature of economic planning in terms of properties observable in reality; (2) to ascribe a meaning to the term which does not depart too radically from common speech; and (3) to avoid, so far as possible, coloring the term with emotional biases which would make it an unsatisfactory tool for logical or empirical investigation.

Two Approaches to Economic Planning. Economic planning may be approached basically in either one of two related yet distinctly different ways. In the first (or "particularized") approach the term connotes a "type" of economy or society characterized by one or a number of particular traits, e.g., "planned economy," "centralized" planning, and so on. In this case, economic planning might be considered a unique and modern phenomenon.³ This approach seems highly dubious.

In a second (or "generic") approach, the term connotes one general characteristic, viz., the attempt to apply reason and foresight to the ordering of human affairs and the attainment of human purposes. From this point of view, a "plan" becomes essentially a program for future action; "planning" becomes the process of preparing programs for future action; and "economic planning" becomes the process of preparing programs for future action in relation to objectives, instruments, and issues in which "economic" aspects, roughly and perhaps arbitrarily defined, play an important role. In this case, economic planning could not be regarded as involving a specific, concrete, or particular type of economy or society, but rather would be looked at as a general characteristic or property, abstracted from the nature of human action and the economizing process; and it could quite easily appear in many and diverse situations, as, e.g., capitalistic planning, socialistic planning, democratic planning, totalitarian planning, business planning, governmental planning, and so on.

The first approach has been most prominent, historically, in a small and highly technical part of the literature; yet it has received the sanction of a large body of planning theorists, and has been used in various forms by such divergent writers as Mises and Hayek, on the one hand, and Sweezy and Dobb, on the other. Thus, Hayek, after some misgivings, defines economic planning as "a central direction of all economic activity according to a single plan, laying down

Economics and the Theory of the State; Dahl and Lindblom, *Politics, Economics and Welfare*; Johr and Singer, *The Role of the Economist as Official Adviser*; Keilhau, *Principles of Private and Public Planning*; Tinbergen, *On the Theory of Economic Policy*; Smithies, et al., *Economics and Public Policy*.

3. Zweig, *The Planning of Free Societies*, apparently thinks this is the case.

how the resources of society should be 'consciously directed' to serve particular ends in a definite way."⁴

Contemporary economic analysis has by no means fully abandoned this older and more particularized approach to economic planning; but there seems to have occurred in recent years a strong shift in emphasis toward a more generic outlook. Dahl and Lindblom, e.g., in their recent study of economic planning, define the term as "an attempt at rational calculation and control in the use of scarce resources."⁵

A possible advantage of the particularized approach is that it may help to restrict one's problems to manageable proportions and to give specific and concrete content to the definition. Another possible advantage is its historical acceptance by a large number of very technical and sophisticated students of the subject.

On the other hand, a striking disadvantage of the first approach is that by thus injecting a particularized content into the definition, it tends to ignore the fact that planning is in reality a complex, many-sided process, with applications in many and diverse surroundings and circumstances. A second possible disadvantage is that it tends to lead to polemical and highly emotionalized debate, and to the construction of stark and unrealistic alternatives, such as planning *versus* freedom, planned economy *versus* the price system, and so on.

A possible disadvantage of the generic approach is that it can mean anything and everything; in short, it may be too general.⁶ This disadvantage, however, would seem to be only apparent. If one defines planning as consisting of anything (within a framework of reason and foresight, etc.) from 1 to n , there is nothing to prevent the identification of, say, "category 162" as the "type of economic planning in mind in this particular study." I.e., a generalized approach may include any number of particularized "types" as subdivisions. If this is the case, then lack of historical "acceptance" of the generic approach within a rather small portion of the planning literature need not be of major importance. Further, the generic approach would seem to correspond more closely to the use of the term in common speech.

4. Hayek, *The Road to Serfdom*, p. 35. See also Hayek, "The Use of Knowledge in Society," in *Individualism and Economic Order*, pp. 78-79; and "The Nature and History of the Debate," in *Collectivist Economic Planning*, pp. 21-23. For a similar statement from a strikingly different source, see Sweezy, *Socialism*, pp. 232 ff.

5. Dahl and Lindblom, *op. cit.*, pp. xxii, 20 ff., 370, 526.

6. Cf., e.g., the criticisms of Hayek, *The Road to Serfdom*, p. 26, and Fisher, *Economic Progress and Social Security*, pp. 190-91.

Whatever the disadvantages of the second approach to economic planning (and there are surely others in addition to those listed here), it has two possible advantages which would seem to make it at least plausible and useful as a supplement to the older approach: First, it recognizes that planning is complex and many-sided, and that adequate one-sentence "definitions" (except generically) of its essential nature are impossible. Second, it forces us, if we are to go beyond general statements about human reason and foresight, away from the "grand alternatives" and the "battle of the 'isms'"⁷ to the really important questions: Who is to plan? How much planning? What types of planning? What are the relations between planners and administrators, between planners and the structure of government? Toward what objectives should planning be directed? Who is to plan the planners? These, and not the comparison of the logic of grandiloquent alternatives, are the real, concrete, and operational questions facing economic planning in democratic societies today.

To sum up, it should be quite apparent that the very concept of economy, economizing, and, indeed, of economic analysis, implies economic planning.⁸ In so far as economic analysis is concerned with human action, based on foresight and reason, it is planning analysis. As Keilhau so effectively phrases it: "Every economy implies a plan. Economy consists in a totality of planned efforts to realize certain purposes. Planning is essential in every activity of economic character. The theory of economic planning does not deal only with some special forms of human societies, but with a certain important part of every economy. It belongs to the general theory of economics."⁹

Economic Planning and Related Concepts. Economic planning in its generic sense is closely intertwined with other concepts, issues, and phenomena, such as policy, administration, calculation and control, and rationality. Indeed, one of the most interesting aspects of the recent shift in emphasis toward a more generalized approach has been a growing rapprochement among welfare economics, the theory of public policy, and the analysis of economic planning, as well as a healthy and growing concern of economists with allied, especially social scientific, disciplines and fields of study.

First, a generic approach to economic planning tends to blur the distinction between "planning" and "policy." If a policy is

7. On this phraseology, cf. Dahl and Lindblom, *op. cit.*, chap. 1.

8. On this point, see the useful article by Bode, "Plan Analysis and Process Analysis," *American Economic Review*, XXXIII (June 1943), 215-41.

9. Keilhau, *op. cit.*, pp. 16-17.

defined as a program of action and a plan as a program for future action, any rigorous distinction between the two in terms of such traditional criteria as the scope or type of economic planning becomes impossible. To many observers, especially supporters of the "planned economy" approach, this may seem highly unfortunate,¹ although, for reasons already given, the advantages of the generic approach would seem to indicate its plausibility and usefulness for certain purposes.

The difference between planning and policy, for our purposes here, lies in terms of the time perspective. A plan is forward-looking; it involves prior preparation. Now, a plan may be selected and executed, in which case it becomes operative policy; or, it may never get beyond the blueprint or planning stage, in a restricted sense. Conversely, a policy may be based upon a considerable degree of prior planning; or, like Topsy, it may "just grow."

Further, policy itself, like planning, may have a variety of meanings. E.g., if the word policy is used to refer to policy "objectives" or to some mixture of objectives and programs of action, then one might argue that policy-making may (perhaps must) precede the formulation of plans.² However policy is defined (and there is little point in pursuing alternative definitions here), its history, performance, and achievement provide a good portion of the data with which economic planning is concerned.

Second, planning cannot be severed rigorously from administration, i.e., the execution of or the organizational and institutional apparatus concerned with policies or plans. To become effective and operative a plan must be translated into administration; the planning process, in a broad sense, includes not only (a) the assessment of conditions and the formulation of alternative plans, and (b) the selection of a particular plan or plans from various alternatives, but also (c) the execution of the plan by an administrative apparatus. As Sir Oliver Franks so effectively phrases it, the first problem is one of intelligence; the second, one of judgment; the third, one of organization and communication: "Successful determination of policy requires high intellectual ability to hold the many diverse factors clearly before the mind and grasp firmly their importance and their relationships. It requires, too, a rare degree of practical judgment to assess rightly the possibilities of the situation and choose what combination of results should be aimed at. And then there are the problems of com-

1. See, e.g., Landauer, *Theory of National Economic Planning*, p. 13, who argues: "A plan should be distinguished from a mere determination of policy." (Note especially the word "mere.")

2. See, e.g., Millett, *The Process and Organization of Government Planning*, chap. 1.

munication. The best decisions are stultified if they are not conveyed to and understood by those who have to carry them out."³

Third, effective economic planning depends on the development and application of methods and techniques of calculation and control. Planning requires a systematic assessment of conditions and comparison of alternatives, a weighing of benefits and costs, a forecasting of future conditions and of the impact of plans upon economic activity and so on. It is this "intelligence" concept or aspect of planning which has been emphasized so heavily by political scientists such as Merriam⁴ and Gaus,⁵ and, more recently, by the political scientist-economist team of Dahl and Lindblom.⁶

Planning also requires methods or social processes of "control." In regard to government economic planning, particularly, control is an unfortunate term. It seems to imply "command," "regulate," and so on, whereas "the hand of government in economic affairs is not always negative or restraining. It can be and often is, positive and promotional." I.e., "government economic control is a two-sided action. It is restrictive, and it is assistive. Any and all controls have something of each characteristic in them."⁷ To become effective, plans must be executed. Execution requires techniques of control. But control in the sense of command is clearly minor in contemporary economic activity. Controls, as aids to more effective planning, are generally much more complex and sophisticated, involving attempts at the manipulation of rewards and punishments, symbols and environments, as well as direct command.⁸

Fourth, economic planning bears an obvious relationship to human reason in general and to the rationality postulate in particular. One must be careful, however, not to become too enamored with the possible equation of economic planning and "rational action."⁹ Even the most systematically and carefully formulated plan can run into snags in execution. Effective planning requires not only social intelligence, but also will and imagination. Thus, the tentative hypothesis

3. Franks, *Central Planning and Control in War and Peace*, p. 27.

4. Merriam, "The Place of Planning," in Harris, ed., *Saving American Capitalism*, pp. 159-60.

5. Gaus, "The Planning Process in Government," in McCormick (ed.), *Problems of the Postwar World*, chap. 7.

6. Dahl and Lindblom, *op. cit.*, chap. 3.

7. Steiner, *Government's Role in Economic Life*, pp. 16-17, 18.

8. On the general problem of controls in relation to economic planning, see Dahl and Lindblom, *op. cit.*, chap. 4; Zweig, *op. cit.*, Part III, chaps. 2-4; Gilbert, "Controls," in Harris, ed., *op. cit.*, pp. 168-74.

9. For an extreme example of the tendency to identify "planning" with "rational action," see Meyerson, et al., *Politics, Planning, and the Public Interest: The Case of Public Housing in Chicago*, Appendix.

is submitted that a generalized theory of economic planning might plausibly and usefully explore *both* the formulation of programs of action, designed to obtain objectives regarded as desirable, and the economic, political, and administrative problems and obstacles which place serious limitations and restrictions upon the attainment of rational economic action in practice.

THE STRUCTURE OF ECONOMIC PLANNING

The adoption of a generic approach to economic planning precludes the possibility of identifying the concept with some particular form or type of economy or society, as has been indicated. A clearer appreciation of some of the specific implications of this approach may be obtained by moving directly to (1) the structure and (2) the process of economic planning.

An exploration of the "structure" of economic planning means essentially "mapping" or "structuring" its basic characteristics, i.e., delineating criteria and indicia in terms of which various types of economic planning might be classified,¹ just as in the case of industrial markets, indicia and criteria are selected for classifying types of market structure, e.g., monopolistic competition, oligopoly, and so on. It is obvious that economic planning is a complex clustering of problems, instruments, issues, criteria, and so on, and that no one listing of relevant indicia will necessarily coincide with another or

1. Among various efforts of this type, see Lewis, *Principles of Economic Planning*, Preface; and Lauterbach, *Economic Security and Individual Freedom: Can We Have Both?* pp. 66-86. Lewis distinguishes six possible types of economic planning: "First, there is an enormous literature in which it refers to the geographical zoning of factories, residential buildings, cinemas, and the like. Sometimes this is called 'town and country planning' . . . Secondly, a very large number of countries have published documents setting out programs of public expenditure over two or four or five, or even ten years. In this literature, 'planning' means only deciding what the government will spend money on in the future . . . Thirdly, in a small, highly specialized part of the literature of economic theory, a 'planned economy' is one in which each production unit (or firm) uses only resources allocated to it by quota, and disposes of the product exclusively to persons or firms indicated to it by central authority . . . Fourthly, . . . 'planning' sometimes means any setting of production targets by the government, whether for private or for public enterprise. . . . (Fifthly) . . . targets are set for the economy as a whole, purporting to allocate all the country's labor, foreign exchange, raw materials, and other resources between the various branches of the economy. . . . And, finally, the word 'planning' is sometimes used to describe the means which the government uses to try to force upon private enterprise the targets which have been previously determined." Lauterbach argues that all possible types of economic planning "boil down" to the following eight major groups: (1) the Fascist type; (2) the Soviet type; (3) Interventionism; (4) Framework Planning; (5) Full Employment Planning; (6) Production-Consumption Planning; (7) Regulatory Planning; (8) International Economic Planning.

attain complete acceptance. Still, it is suggested that the following criteria, among others, may be of some significance:

The Sphere of Human Action. Economic planning may take place in various spheres of human action, as, e.g., (1) individual, (2) group, and (3) governmental planning. A source of considerable confusion in planning literature is the tendency to equate planning in general with governmental planning in particular. This is unfortunate, for two reasons: First, it tends to imply that planning is somehow a peculiarly political or governmental phenomenon, whereas it has been argued that the very nature of the word economize necessarily implies planning. Second, it implies that a reasonably sharp distinction may be drawn between private and public, governmental and nongovernmental, whereas it would appear more realistic to presume the opposite, viz., that these distinctions are becoming increasingly difficult to make.²

The Geographical or Political Area. Economic planning may relate to various geographical or political areas, as, e.g., (1) regional, (2) national, and (3) international planning.³ Many models of economic planning seem to be based on the rather tenuous presumption that planning somehow is not "really planning" unless it takes place on the national level, for the national economy as a whole. This tends to overlook the very interesting planning work being done in other geographical or political areas.

The Scope of Economic Planning. Economic planning in practice, if not always in theory, is by no means coterminous with "total" or "comprehensive" planning, or with "planned economy." The scope of economic planning can and does vary all the way from the relatively limited planning of an essentially private or mixed enterprise economy to the comprehensive, all-embracing planning of a totalitarian economic system. A federal budget, for example, is a plan, and the budgetary process is an important aspect of, as well as example of, economic planning in the American economy.⁴ The procedure of equating planning with total planning has two unfortunate consequences: First, it raises the unresolvable issue of determining: How total is "total planning?"⁵ Second, if true planning must be compre-

2. See, e.g., Clark, "America's Changing Capitalism: The Interplay of Politics and Economics," in Berger (ed.), *Freedom and Control in Modern Society*, pp. 192-205.

3. See, e.g., Zweig, *op. cit.*, Part III, chap. 5.

4. Smithies, *The Budgetary Process in the United States*, pp. 21 ff., uses somewhat different phraseology, but comes to essentially the same conclusion.

5. As an example of how a categorical equation of planning and planned economy can put the concept of economic planning into an intellectual strait-jacket in which even strong critics of planning find difficulty maneuvering, see

hensive," then the concept and analysis of economic planning has little, if any, application to empirical problems of public policy in Western, industrially advanced, liberal democracies, none of which have embarked or are likely to embark in the near future upon programs of total or comprehensive economic planning.

The Objectives of Economic Planning. Planning activity is purposive activity. It is difficult to think of planning in any concrete and specific sense except as related to purposes and objectives. Planning, in a sense, may be referred to as a neutral instrument, applicable to any objectives. Yet, it is also true that the nature of economic planning is structured by the objectives toward which it is directed. Defense planning, e.g., is not full-employment planning, and so on. One of the major limitations of planning theories in the 1920's and 1930's was the overemphasis placed upon "rational resource allocation" as a planning objective. From an operational standpoint, the relevance of this goal, at least as an all-embracing, monistic purpose, seems rather limited. Soviet planning, e.g., is more concerned with economic development and military preparedness than with allocational efficiency in the traditional neoclassical sense, while democratic socialists seem to be more interested in such objectives as full employment, economic stability, security, and various broadly defined "social welfare" measures.⁶

Economic versus Noneconomic Planning. The distinction between economic and noneconomic planning is a tenuous one at best. If one were to take the resource allocation definition of economics seriously, then any human effort to economize in the allocation of scarce resources, for any purpose, in any sphere of action, could be regarded as economic planning, in which case, politics, e.g., would be more or

Hayek, *The Road to Serfdom*. At one point (p. 35), Hayek identifies planning with "a central direction of all economic activity according to a single plan . . ." and so on, as noted above. At another point, however (p. 105), it is maintained that "once the free working of the market is impeded beyond a certain degree, the planner will be forced to extend his controls until they become all-comprehensive." But if this is so, then it logically follows that (at least temporarily) "planning" may exist which is not "truly planning" in an "all-comprehensive" and total sense.

6. For cogent criticisms of contemporary planning theory in terms of the relevance of the objectives postulated, see Baran, "National Economic Planning," in Haley (ed.), *A Survey of Contemporary Economics*, II, esp. 383-84; Sweezy, *Socialism*, esp. pp. 232-38; Dobb, "A Review of the Discussion concerning Economic Calculation in a Socialist Economy," in *On Economic Theory and Socialism*, pp. 55-92. For a systematic recognition and analysis of the multiplicity of planning objectives, see Smithies, "Economic Welfare and Policy," in Brookings, *Economics and Public Policy*, chap. 1; Clark, "Economic Welfare in a Free Society," in Lekachman (ed.), *National Policy for Economic Welfare at Home and Abroad*, chap. 7; and Millikan, "Objectives of Economic Policy in a Democracy," in Millikan (ed.), *Income Stabilization for a Developing Democracy*, chap. 1.

less swallowed up by economics.⁷ If, on the other hand, the essence of the "economic way of thinking" is *individual* choice,⁸ then it would appear that a good portion of group and collective action must lie outside the field of economic planning.

Perhaps the most empirically useful, yet methodologically frustrating, solution to this problem is simply to acknowledge that clear-cut distinctions between economic and social, etc., planning, and between economic and political aspects of planning, are difficult, if not impossible, to make, yet to continue to identify economic planning, roughly speaking, with the preparation of programs in relation to goals, measures, and issues in which "economic" aspects, loosely and perhaps arbitrarily defined, play a significant role.

Economic Planning and Social Processes. An interesting aspect of recent economic and political theory has been the attempt to relate the planning concept to underlying social processes, drawn upon as means of economizing in resource allocation and in attaining various economic and noneconomic objectives. Dahl and Lindblom, e.g., in their recent study, have distinguished four social processes, each demonstrating relations between leaders and nonleaders:⁹ (1) the price system (control of and by leaders); (2) hierarchy, e.g., bureaucracy (control by leaders); (3) "polyarchy," or democracy (control over leaders); and (4) bargaining (control among leaders). A potentially fruitful result of this typology is that the price system, instead of being regarded as the antithesis of economic planning is itself understood as a social process essential for planning. Although no society could use all four of these processes simultaneously to resolve all human problems, a society can conceivably use some combination of all four in the resolution of the issues of economic planning.

Public Ownership. Economic planning may conceivably take place under varying conditions of public ownership or control of industry. The traditional assumption of the "planned economy" definitions of economic planning¹ has been that planning and government ownership are closely intertwined, that, in short, either

7. Cf., e.g., Frank Knight, "Ethics and the Economic Interpretation," in *The Ethics of Competition*, p. 34: "In so far as the ends are viewed as given, as data, then all activity is economic. . . . From this point of view, the problem of life becomes simply the economic problem . . ." This aspect of the resource allocation definition plays a major part in contemporary criticisms of "planned economy." See, e.g., Hayek, *op. cit.*, pp. 89 ff.

8. Cf. Ellis, "The Economic Way of Thinking," *American Economic Review*, LX (Mar. 1950), 1-12. But compare Allen, "Is Group Choice a Part of Economics?" this *Journal*, LXVII (Aug. 1953), 362-79.

9. Dahl and Lindblom, *op. cit.*, pp. 171 ff.

1. Cf., e.g., Dickinson, *The Economics of Socialism*, pp. 2-29.

"unplanned socialism" or "planned capitalism," "although logically thinkable," are "unlikely to occur in practice."² Recent experiences in planning and nationalization in Scandinavia and England, however, would seem to indicate the necessity for greater flexibility on this score. Certainly among democratic socialists, the idea of public ownership as a necessary prerequisite for effective planning is declining in significance,³ while the successful execution of governmental economic plans in any institutional framework requires a considerable amount of persuasion, inducement, and regulation, quite apart from, and in addition to, public ownership. The growing recognition that economic planning is not necessarily tied up with any particular set of ownership or property relations has been accompanied, as is well known, by a growing literature on the planning of "mixed economies."⁴

The State of Economic Development. Economic planning may take place under various types, stages, and states of economic backwardness or advancement, delineated both quantitatively (e.g., real income per capita) and qualitatively (e.g., the maturity and development of the banking, transportation, and communications systems, the diversity of the output). If one were interested in being schematic, it might be possible to distinguish among economic planning in relation to (1) establishing requisite technological-institutional conditions for economic development; (2) fostering more rapid development in relatively underdeveloped areas; and (3) maintaining economic growth in highly developed countries. Whatever typology one applies, it would seem clear that the nature and state of economic development is becoming an increasingly important criterion and issue in contemporary economic planning.

Centralization versus Decentralization. Economic planning may be structured by various procedures and techniques of private and public organization and administration, e.g., the degree of centralization versus decentralization, i.e., (1) the degree of localization versus centralization of decision-making, or the extent of local autonomy versus central control; and (2) the degree of separation versus concentration of powers, or the extent to which one agency or institution, say a legislature, is supreme versus the extent of checks and balances involved in a division of powers. It has sometimes been maintained that planning is not "really" planning unless it is centralized, i.e., unless planning measures emanate from a supreme, central governing

2. *Ibid.*, p. 16.

3. Cf., e.g., Bjerve, "Government Economic Planning and Control," in H. Friis (ed.), *Scandinavia Between East and West*, pp. 52 ff.

4. Cf., e.g., Meade, *Planning and the Price Mechanism*; Clark, *Alternative to Serfdom*; Suranyi-Unger, *Private Enterprise and Governmental Planning*.

authority, and are concerned with the economy as a whole.⁵ But, clearly, economic planning can and does involve various combinations of centralization and decentralization, in either or both of the two above-mentioned senses. Referring to governmental administration, for example, Leslie Lipson distinguishes four possibilities: (a) separation of powers and decentralization or localization (e.g., despite increasing federal controls, the United States); (b) concentration of powers and decentralization (e.g., Switzerland); (c) concentration of powers and centralization (e.g., France); (d) separation, yet centralization (e.g., Costa Rica).⁶

Democratic versus Nondemocratic Planning. Unless one is a complete dogmatist or resolves the issue by definition, one recognizes the possibility of economic planning under various political systems, e.g., democratic or nondemocratic. The relative neglect of this criterion in the past has led more than one observer to draw parallels (which do exist) between capitalistic controls in wartime and planning under Fascism (using the term generically), and between different varieties (democratic and nondemocratic) of "socialistic" planning measures. Conversely, one fruitful result of including this criterion in an analysis of the structure of economic planning is that it helps to emphasize basic similarities between "capitalistic" and "socialistic," yet democratic, planning, and between different varieties of totalitarian planning measures, whether Fascistic or Communistic.⁷

Institutional versus Operative Planning. Lorwin⁸ has distinguished between "institutional" and "operative" planning, i.e., between the formulation of a framework of institutions within which individuals are left relatively free to plan (e.g., antitrust policy, monetary and fiscal policy, social security programs) and the specific supervision and direction of economic activity. This distinction is not completely

5. Cf., e.g., the writings of Hayek and Dickinson, noted above. See also Halm, *Economic Systems: A Comparative Analysis*, pp. 191-92.

6. Lipson, *The Great Issues of Politics: An Introduction to Political Science*, pp. 268 ff. Incidentally, Lipson observes (p. 270n) that all four countries cited above "are democracies. Presumably, therefore, democracy may co-exist with any of these four patterns." This is an elemental, perhaps even an elementary, point, but its emphasis helps us to appreciate the superficiality of the argument which maintains that "central" planning must, by definition, constitute a threat to democratic institutions.

7. On democratic planning, and the necessity of democratic as opposed to totalitarian planning, see Merriam, "The Place of Planning," in Harris (ed.), *op. cit.*, p. 161, and *The New Democracy and the New Despotism*. On the political characteristics of totalitarianism and their impact on economic planning, see, among others, Friedrich and Brzezinski, *Totalitarian Dictatorship and Autocracy*. See also Halm, *op. cit.*, pp. 192-97, for an interesting application of this criterion to the classification of types of economic planning and economic systems.

8. Lorwin, *Time for Planning*, pp. 13-15.

clear, and, of course, the two types of planning tend to overlap in practice. Further, quite understandable differences of opinion exist as to the precise content of institutional planning.⁹ Still, the distinction, although rough, may be useful, for, as Nourse remarks: "Democratic national planning for a high level of well-coordinated economic activity with operation largely in private hands is quite different from a government plan supplemented by government operation."¹

This list of criteria of the structure of economic planning is by no means complete, nor are the criteria completely isolated from one another. It could be extended or modified. If desired, interrelations among various criteria could be explored in more detail and formal classifications of types of economic planning could be constructed. The purposes of the above listing, however, were much less ambitious. The point has been simply to indicate, in a general way, the complexity and many-sided nature of economic planning, and some of the more relevant criteria involved in any attempt at mapping its structure.

THE PROCESS OF ECONOMIC PLANNING

Even the most intricate map of the structure of economic planning, however, and the most sophisticated of typologies, would be artificial and incomplete, if they were not supplemented by a recognition that economic planning is a dynamic "process," i.e., by a recognition of (1) the nature and significance of the time factor or time dimension so crucial to the concept of economic planning; (2) the "steps" or "facets" of economic planning; and (3) the possibility of the transformation of economic planning itself and of the surrounding economic and political system in which it operates and unfolds.

Economic Planning and the Time Dimension. It is obvious that the concept of economic planning injects the time factor and the problem of process into the center of economic analysis. The essence of economic planning is that it is futuristic; it is forward looking; it involves systematic thought and preparation *ex ante*; it involves "pre-time"² thinking, i.e., thinking prior to the unfolding of a sequence of events. Needless to say, the remarks here will not constitute an intensive analysis of the role of time in the planning process,³ much

9. E.g., most of the policy recommendations of both Henry Simons, *Economic Policy for a Free Society*, and Keynes, *The General Theory of Employment, Interest and Money*, chap. 24, could be included under the heading of "institutional" planning.

1. Nourse, "Serfdom, Utopia, or Democratic Opportunity?" *Public Administration Review*, VI (Spring 1946), 185-86.

2. Keilhau, *op. cit.*, p. 44.

3. *Ibid.* See also Millett, *op. cit.*, pp. 43-54.

less the application of the planning concept to dynamic economics or expectational problems.⁴ However, a few elemental observations are necessary to indicate the range of issues involved and the importance of the time dimension.

First in a series of problems raised by recognition of the time dimension is the time perspective involved, e.g., short-range versus long-range planning. This distinction, however, in economic planning no less than in the theory of the firm, is a tenuous one at best, since a long-range program may be the summation of a series of short-range programs, while long-range programs provide the framework within which short-range plans are developed.

On the one hand, it is possible to overemphasize long-range projects. In Soviet planning, e.g., Gosplan has been divided, since the Twentieth Party Congress, into two divisions, one responsible for long-range programming, the other in charge of short-run projects. Apparently this administrative alteration was an attempt to counter the tendency of Gosplan officials and researchers to neglect short-run projects for long-range planning. Similarly, in American planning, to take a second example, the National Resources Planning Board was often criticized for its heavy emphasis on long-range issues which, according to Millett, "at times amounted almost to a psychopathic desire to escape reality"⁵ and to avoid recommendations.

On the other hand, it is possible to become overly concerned with resolving current issues, to the neglect of the longer-run implications of short-run programs.⁶ One of the greatest assets and contributions of the economist, it would seem, is, or at least ought to be, the capacity for maintaining an extended or long-range time perspective, and recognizing the impact of short-range programs upon longer-run conditions and objectives.

A second problem is that the preparation of plans themselves is a time-consuming process. If plans are to be available to meet contingencies, they must be prepared in advance. On the other hand, if plans are prepared too far in advance, altered conditions may make them inapplicable at the time of execution. Fortunately, in some spheres of planning this possible conflict may be resolved, or at least diminished, as, e.g., in the case of public works planning in the United

4. Cf. Lindahl, *Studies in the Theory of Money and Capital*, Part I; Hart, *Anticipations, Uncertainty, and Dynamic Planning*; Svernilson, *Ekonomisk Planering*; Shackle, *Expectation in Economics*; Hicks, *Value and Capital*, Part IV.

5. Millett, *op. cit.*, pp. 52-53.

6. Smithies, "Economic Welfare and Policy," in *op. cit.*, p. 21; Viner, "The Short View and the Long View in Economic Policy," *American Economic Review*, XXX (Mar. 1940), 1-15.

States. Prior to World War II, appropriations for public works planning were coterminous with appropriations for the construction of projects themselves. Since 1940, however, planning appropriations have been made in advance, the principle of a "shelf of public works" has become firmly entrenched in American planning, and public works agencies present programs several years in advance.

Closely related to this issue is the problem of timing and time lags. The importance of timing is often emphasized in planning literature. In a rapidly changing economy, an ill-timed application of a planning measure may be worse than no application at all. The timing problem becomes especially crucial in view of so-called "lags in response." Milton Friedman has divided "into three parts the total lag involved in any action to offset a disturbance: (1) the lag between the need for action and the recognition of this need; (2) the lag between recognition of the need for action and the taking of action; and (3) the lag between the action and its effect."⁷

A third problem is the uncertainty of the future. As Keilhau phrases it, the act of looking forward by the planner requires a recognition of two types of "future." "In the first instance we have all those happenings which are going to take place completely independent of our own will. We choose to refer to this immense totality as the *conjunctural future* . . . but secondly we have before us a limited number of events which must or may be influenced by our personal resolutions. We will call it the *affectable future*. The former represents the sphere of prediction and prognosis. The latter represents the sphere of planning."⁸ Two important, although perhaps obvious, observations seem relevant here: (1) The extent to which the future is "affectable" by economic planning depends, among other things, upon the scope of planning. Other things remaining equal, the "conjunctural future" would seem to be of greater relevance under limited than under comprehensive planning. (2) The future, conjunctural or affectable, is uncertain, although the degree of uncertainty would seem to be more intense the greater the importance of the conjunctural component of future conditions. What this boils down to, in programmatic terms, is that limited planning must often be based upon inadequate analysis and prognosis,⁹ and must be content with relatively limited and restricted ambitions.¹

7. Friedman, "A Monetary and Fiscal Framework for Economic Stability," *American Economic Review*, XXXVIII (June 1948), 245-64; reprinted in *Essays in Positive Economics*, pp. 133-56. The above quotation is found on p. 145.

8. Keilhau, *op. cit.*, p. 48.

9. Cf., e.g., Smithies, et al., "National and International Measures for Full Employment," pp. 81-82.

1. Cf., e.g., Popper, *The Open Society and its Enemies*, pp. 154-64.

Fourth, there is the problem of planning periods. The large bulk of planning theory of the formalized type in economic literature has been of the period analysis variety, i.e., has employed the techniques of "discrete processes" and difference equations. Erik Lindahl, for example, a leading exponent of this approach, has distinguished (1) the "external conditions" impinging upon economic planning; (2) the construction of plans at the beginning of time periods, oriented toward the attainment of certain aims, and based on (a) the assessment of past developments, present conditions, and expectations regarding future conditions and developments, and (b) the choice of a plan from alternative possible courses of action; and (3) the dynamic process of the revision and alteration of plans with the passage of time, as new information becomes available and the results of past action become clearer.²

Quite apart from the formal structure and details of an analysis of this type, which can be quite an imposing and complicated affair, it may be relevant, for our purposes here, to note certain broad advantages and disadvantages of the methodological technique of separating the passage of time into discrete periods. First, as Lindahl argues effectively,³ human action does not and cannot operate in terms of a completely continuous recognition and assessment of change in external conditions, internal variables, objectives, and courses of action. The recognition of change typically takes place "intermittently." Hence the notions of time periods or "periods of registration" and so on may have considerable usefulness. Governmental budgets, prepared annually, may provide a rough illustration of this principle. Second, relevant statistical data for national economic planning do not refer to rates of flow, but to amounts during a past period of time.

On the other hand, a formalized model, based on the assumption that plans are not revised during a planning period, first, faces the difficulty that the planning period may be a different length in different sectors of the economy, and, indeed, in different governmental agencies and departments; and, second, may conceal the fact that a general plan cannot lay down intimate directives for all specific problems that may arise during a planning period, and that considerable leeway may be given to administrative officials regarding the timing and content of programs, in terms of their judgment of the altered nature of conditions, *within* a time period in which over-all objectives and expectations remain roughly stable.

2. Lindahl, *op. cit.*, especially pp. 35 ff.

3. *Ibid.*, p. 42.

Facets of the Planning Process. Many current studies of the dynamics of economic planning have grappled with the problem of demarcating its facets, or, more ambitiously, its "steps" or "stages."⁴ Of course, breaking down a complex process into steps, like dividing a complex structure into parts, faces dangers of oversimplification and rigid schematization. Without attempting, therefore, to claim any absolute merit or pre-emptive worth for the approach adopted here, it is suggested that the process of economic planning may be divided roughly into three related, yet not identical, facets: (1) the formulation of objectives; (2) the assessment of conditions and the formulation of alternative plans, based upon expectations of possible future conditions and possible future changes in relevant variables; (3) the formulation and execution of programs of action.

Three tangential remarks may be made: First, this breakdown is a generalized one: It is intended to apply to planning (and closely related activities) in any sphere of human action or endeavor, e.g., governmental and nongovernmental, democratic and totalitarian, limited and comprehensive, and so on.

Second, the term facets, rather than steps or stages of economic planning is used deliberately, thus leaving open the question of the "order" involved in terms of time sequence or analytical interpretation. Various observers have offered reasons why this or that step should be placed first.⁵ But as Jöhr and Singer argue so effectively in their recent study on the methodology of public economic policy, the order of these steps may follow alternative patterns, and is not necessarily restricted to a rigid one-two-three sequence.⁶

Third, the three facets of economic planning itemized above are handy phrases used here to summarize in a few words what in actuality can be a many-sided and highly complicated affair: Each facet could be broken down into as many subdivisions as seemed desirable

4. Millett, *op. cit.*, chap. 2, distinguishes among "the determination of objectives, the measurement of an existing situation, and the design of a program for positive action." Landauer, *op. cit.*, chaps. II and III, distinguishes simply between the "making" of the plan and its "execution." Galloway, *Planning for America*, p. 6, presents a five-step process: "1. The determination of objectives to be sought. 2. Research — to understand the problem. 3. The discovery of alternative solutions. 4. Policy-making — choosing between alternatives, including the frequent choice of doing nothing. 5. The detailed execution of the chosen alternative . . ." Lindahl, *op. cit.*, p. 48, summarizes four "different links in the chain of events" as: "(1) the actual development of the factors relevant to the planner, (2) the latter's view of this development, (3) the resulting changes in his anticipations of future conditions which are of importance for his actions, and (4) the resulting alterations of his plans of action for the nearest period ahead."

5. Cf., e.g., Millett, *op. cit.*, pp. 11 ff.

6. Jöhr and Singer, *op. cit.*, p. 27.

for various purposes. Consider, e.g., the "formulation of objectives." Goals of economic life are essentially an ethical problem, bringing the study of economic planning into interconnection with some of the broadest issues and problems of social philosophy. But even on the more technically "scientific" (here explanatory-predictive) plane of analysis, there are the problems of (1) the sources or origins of planning objectives (example: who maintains that "stabilization" is desirable? individuals? groups? governments?); (2) the nature and content of these objectives (example: how full is "full employment"?); (3) the degree of compatibility or incompatibility among competing objectives (example: to what extent is a competitive market structure, in the sense of the dispersion of economic power, compatible with technological progress and dynamic performance?).

Similarly, a more detailed analysis than is possible here might involve the dissection of the "assessment of conditions" and the "formulation of plans" into their constituent parts.⁷ Regarding the assessment of conditions, at least four subdivisions might be distinguished: (1) investigation of the existing situation (situation here referring possibly to valuational or institutional characteristics of the particular economic system under consideration as well as such traditional economic variables as incomes, prices, employment, and so on); (2) ascertainment of the extent to which existing conditions differ from those conditions necessary for the attainment of objectives — i.e., the confrontation of values with facts; (3) the ascertainment and assessment of the effectiveness or success of past policy measures in relation to past aims and conditions; (4) the forecasting of relevant future conditions (conditions here again referring possibly to future changes in objectives and surrounding socio-economic institutions as well as to possible changes in more traditional economic variables).

In similar fashion, the formulation of plans might involve (1) the construction of possible alternative measures (and combinations of measures) to attain objectives; (2) the assessment of the possible success of alternative measures in attaining objectives, under alternative possible conditions, which implies (3) forecasting the effects of the measures, both upon the objectives postulated and upon other values, programs, institutions, and variables.

The "formulation and execution of programs of action," like the other two facets listed, can also be broken down into subcomponents, e.g., (1) the selection of a plan from various alternatives; (2) obtaining support or ratification of a plan; (3) formulating and developing the implications of a general plan or program in specific terms; (4) execut-

7. Cf., e.g., *ibid.*, Part I, chaps. 4-6.

ing the plan; (5) reviewing the execution of plans.⁸ One might also want to mention (6) "planning the planners,"⁹ i.e., formulating and executing plans for co-ordinating and checking planning (and related administrative) personnel.

Economic Planning and the Alteration of Economic Systems. One of the most fascinating, and undoubtedly unresolvable, issues injected into the discussion of economic planning by a recognition of the dynamics of the planning process is the possibility of alteration in the structure of the process itself, and of the politico-economic system of which it is a fundamental component. Possible cases and examples of relations between economic planning and institutional change are numerous. Two possibilities are selected for very brief comment here.

One possibility is the one which was nearly discussed to death in the early postwar period, viz., the "Road to Serfdom" controversy.¹ This debate was unfortunately clogged with a series of methodological, definitional, and emotional obstacles which made effective discussion highly difficult. However, clearing away this underbrush as best we can, we are left with two central questions: (1) Is it likely that contemporary programs of economic planning in Western societies will become "total and comprehensive?" (2) If this happened, could or would pressures for the imposition of totalitarian controls be avoided?

Both questions, of course, are speculative and contingent, i.e., are futuristically oriented and can be answered completely only as the course of historical development unfolds. Further, although both questions are speculative, the second is more remote, since it is based upon the elimination of the first. The two questions are related, however, because of the possibility of the erosion of freedoms prior to the attainment of comprehensive planning.

This does not mean that we must wait for history to give us the answers, or that we can say nothing about these issues. It does imply, however, that we lack empirical evidence of a conclusive sort, drawn from past experiences, and that the future structure of economic systems and the character of economic institutions under varying types and degrees of planning is not foreseeable via the tools of logic. Stark alternatives, in logical terms, display a disconcerting habit of becoming workable mixtures in practical action.

8. Cf., e.g., Smithies, *The Budgetary Process in the United States*, chap. II.

9. Wooton, *Freedom under Planning*, chap. 10; Mannheim, *Man and Society in an Age of Reconstruction*, p. 74.

1. Cf. Hayek, *The Road to Serfdom*; Finer, *The Road to Reaction*; Wooton, *op. cit.* Cf., also, Clark, *Alternative to Serfdom*; Jewkes, *Ordeal by Planning*; Wright, *Democracy and Progress*; Meade, *Planning and the Price Mechanism*.

What we can do is to formulate a qualitative assessment of the general direction and momentum of social movements, and to anticipate, in a broad way, some of the problems which might arise, by judicious application of such historical and institutional materials as we have that seem relevant. Sweeping generalizations and rigid conclusions must be regarded with extreme caution in this area. Against the categorical vision of "creeping socialism," compare, e.g., the guarded judgment of J. M. Clark: "Some observers appear to hold, or at least imply, that once the change from individualism starts, it can find no logical point of rest short of complete collectivism. As against this, I regard it as more probable (though naturally unprovable) that a society of Western type which has once thoroughly tasted the spirit that combines democracy with economic freedom and has gained prosperity under it would check any trend toward collectivism without reaching a point of no return. This would mean that its natural state is one of uneasy balance between private and community action."²

As a second example, it is interesting to note that students of the Soviet Union have been pushing toward an analysis of economic planning and transformation of politico-economic systems from an entirely different angle. Soviet planning, so the argument runs, has been directed basically toward two objectives: (1) to control the economy, as a means of increasing political power and ensuring political loyalty, i.e., planning as an instrument of totalitarian control; (2) to plan economic activity so as to increase productivity and the economic base of the country, i.e., planning as an instrument of economic growth.

The first objective has required the familiar apparatus of totalitarian dictatorship. The second objective, however, has necessitated the development of an industrial bureaucracy, the placing of members of the administration in positions of responsibility and power, and the admission of the administrative and technical elite into the Party. The threat of these developments for the maintenance of totalitarian political control, it is maintained, is direct: The efficient exercise of responsibility by a bureaucracy requires rationality, order, stability, hierarchy, chains of command, rules regarding the allocation of authority, and so on. But these are characteristics of authoritarianism,³ not totalitarianism. The problem is intensified because of

2. Clark, "Economic Welfare in a Free Society," in Lekachman (ed.), *op. cit.*, pp. 295-96.

3. Cf., e.g., Moore, *Terror and Progress in USSR*, p. 19, who argues: "If the Communist rulers ever achieved the degree of regularity and precision in their society that they seek so constantly and vigorously, they would seriously

the impossibility of severing rigorously the determination of objectives and the selection of policies from the execution of plans and administrative decisions. The Party leaders are constantly faced with the possibility of the technical and administrative intelligentsia usurping their power.

Thus far, at any rate, a rough balance has been maintained in Soviet planning between the political requirements of totalitarian control and the requirements of economic development and bureaucratic-administrative stability and hierarchy. However, if, in the future, the requirements of economic development (especially in agriculture) and bureaucratic efficiency do emerge triumphant over the objective of political control, the rationality and regularity involved in the bureaucratic outlook may conceivably alter the bases of totalitarian dictatorship in Soviet Russia.

What this highly speculative type of analysis boils down to, for our purposes here, is that perhaps a highly rationalistic, bureaucratically organized system of comprehensive economic planning can function most effectively neither under constitutional democracy nor under totalitarian dictatorship, but rather under some sort of constitutional or authoritarian (but not totalitarian) dictatorship, or benevolent, paternalistic despotism. However, even this is not certain, and the above discussion was mainly for the purpose of definitional illustration, not of detailed analysis.

SUMMARY AND CONCLUSIONS

1. Economic planning lies at the core of the economizing process. It is a potentially useful concept in economic analysis and research, and is currently being reconsidered and re-examined by various observers.

2. Whereas the particularized approach to the definition of economic planning has been accompanied, in the past, by the construction of stark and artificial alternatives, the generic approach leads directly to the "structure" and "process" of economic planning.

3. Economic planning may be structured in various ways, by such factors as the state of economic development, the scope of governmental action, the nature of the political system, and so on. A clear recognition of relevant criteria and indicia of economic planning is a first step toward the construction of more empirically applicable "types" of planning, and models of economic analysis based on these types.

alter and probably destroy the basis of their own rule." On these issues, see also Arendt, *The Origins of Totalitarianism*, Part III, and Rostow, *The Dynamics of Soviet Society*, pp. 191-96.

4. The recognition of economic planning as a dynamic process illustrates the importance of the time dimension in economic analysis, raises the issue of the facets or steps of economic planning, and re-emphasizes the fact that changes in economic planning may bring changes in surrounding socio-political institutions and economic systems.

JOHN E. ELLIOTT.

UNIVERSITY OF SOUTHERN CALIFORNIA

ON THE THEORY OF NEGOTIATION

By CARL M. STEVENS

I. Introduction, 77. — II. Distinction between "bargaining" and "negotiation," 78. — III. The choice theory model used in this analysis, 79. — IV. An explanation of why the parties negotiate at all, 82. — V. Analysis of the negotiation process itself, 85. — VI. An illustrative application of the model, 87. — VII. Five analytically separable ways in which negotiation may terminate, 90. — VIII. Conclusion, 96.

I

In recent literature, increasing attention has been paid to the theory of bargaining and negotiation.¹ This is a most desirable development since, in the light of the number and importance of the economic and other transactions (and interactions) thus mediated, there can be little doubt of the real need for a good theory of the negotiation process.

Interesting as have been some of the contributions to particular aspects of negotiation theory, no general statement of such a theory has yet emerged.² As a contribution to the development of such a general theory, I herein suggest the adaptation of a psychological choice theory model to the analysis of negotiation.

Throughout this paper, collective bargaining over terms and conditions of employment is used as an example of the negotiation process. However, the format of the analysis is intended to be applicable to negotiations of many sorts.³

Emphasis is upon developing the analytical framework rather than upon application of this framework to particular problems. The major points are:

1. Perhaps the most comprehensive recent effort is Neil W. Chamberlain's *A General Theory of Economic Process* (New York, 1955). For an earlier quite comprehensive formulation see G. L. S. Shackle, *Expectation in Economics* (Cambridge, England, 1949), chap. VI, "A Theory of the Bargaining Process." See Harvey M. Wagner ("A Unified Treatment of Bargaining Theory," *Southern Economic Journal*, XXIII, April 1957, 380-97) for a recent discussion of three approaches (bilateral monopoly, risk evaluation, and the theory of games) to bargaining theory. See also J. Pen, "A General Theory of Bargaining," *American Economic Review*, XLII (Mar. 1952), pp. 24-42.

2. See, for example, T. C. Schelling, "An Essay on Bargaining," *American Economic Review*, XLVI (June 1956), pp. 281-306.

3. Including, for example, not only negotiation which is part of "economic" relations in the usual sense of this term, but also negotiation which may be part of intranational and international political relations, etc.

(1) A reasonably self-contained description of the choice theory model to be employed in the analysis.

(2) An explanation, in terms of the model, of why the parties negotiate at all, i.e., rather than selecting some alternative method for mediating the transaction in question.

(3) An analysis, in terms of the model, of the negotiation process itself, i.e., the significance of the information exchanged in the course of negotiation.

(4) An illustrative application of the model to the disputed problem of the "real" function of the so-called basic criteria used in wage negotiations (e.g., comparative wage rates, ability to pay, etc.).

(5) An explanation of the five analytically separable ways in which negotiation may terminate. (It develops that, from this point of view, there is an important asymmetry in the effects of different negotiation tactics).

The choice theory model employed is of the stable equilibrium "conflict" choice type, basically different from models usually employed in the analysis of economic choice behavior. Although herein applied only to negotiation theory, this model may well be of much more general interest to students of economic choice behavior. It may also be noted that since this is an equilibrium model, it permits, analogously to comparative statics analysis generally, prediction of the direction of shift in the variable of interest (e.g., wage rate), consequent, *ceteris paribus*, upon shifts in certain environmental parameters of interest. In this sense then, it makes the outcome of negotiations theoretically determinate.

II

In much of the literature the terms "bargaining" and "negotiation" are used more or less interchangeably, as if they had reference to the same phenomena. However, it is helpful to maintain a distinction between the two. In any exchange transaction, for example, an ordinary retail purchase, a bargain (regarding the terms of exchange) is struck, and hence, a kind of bargaining may be said to have taken place. But, as in this instance, there need be no negotiation involved. Only certain exchange transactions are featured by negotiation, although all may be viewed as instances of bargaining.⁴ Thus negotiation is just one of several ways in which a bargain may be concluded. In order to mediate any transaction, *A* and *B* must

4. On this point see Chamberlain, *op. cit.*, especially chaps. 6, 7, and 8. As Chamberlain illustrates, the concept "bargaining power" may be elaborated with respect to numerous transactions in which no negotiation takes place.

exchange some minimum amount of information, namely the initial terms upon which either (or both) is willing to conclude the transaction (on a take-it-or-leave-it basis), and subsequent acceptance or rejection of the offered terms.⁵ However, *A* and *B* may be said to negotiate if they exchange information relevant to some prospective transaction in addition to this minimum amount. I subsequently discuss the kinds and functions of such additional information.

III

Let us now consider the choice theory model used in the subsequent analysis. It has been observed that a subject, when confronted with a choice between goal objects, may behave in either of two distinctly different ways:

(1) The subject may immediately elect one or the other. Call this a nonconflict choice.

(2) The subject may immediately elect neither goal. Rather he will remain uncertain for a period of time in a sort of behavioral equilibrium between the two. Call this a conflict choice.

How is this difference between nonconflict and conflict choice behavior to be explained?⁶

The model here described bases the explanation of behavioral equilibrium in the conflict choice situation upon the concepts of the "approach gradient" and the "avoidance gradient."⁷ Just as an individual may learn approach tendencies to rewarding (positive) goals, so also he may learn avoidance tendencies to goals such that to elect them involves an expectation of punishment or nonreward (negative goals). The approach gradient is a name given the hypothesis that the strength of an individual's tendency to approach a posi-

5. This exchange of information need not be of the vis-à-vis variety nor need it involve verbalization.

6. Economic choice theory might suggest this explanation: in the conflict choice situation the objects are equally preferred or indifferent, whereas in the nonconflict choice situation the relative preference for one is much greater than that for the other. This explanation is not, however, on any simple interpretation of it, adequate. Psychological choice theory suggests that all nonconflict choices will be made quickly, the matter of relative preference being manifest operationally in the probability that one rather than the other will be elected. Thus, in the case of a nonconflict choice involving two goals of equal preference, we should expect, on a series of trials, that each would be selected immediately about half of the time. (This is, indeed, the operational definition usually given the concept of indifference in experimentation associated with economic choice theory.) The conflict choice situation is of a generically different sort.

7. John Dollard and Neal E. Miller develop the model I shall here describe. See their *Personality and Psychotherapy* (New York, 1950), especially chap. XXII. Dollard and Miller do not apply their model to the negotiation problem.

tive goal is a decreasing function of his distance from the goal. Analogously, the avoidance gradient is a name given the hypothesis that the strength of an individual's tendency to avoid a negative goal is a decreasing function of his distance from the goal.⁸ A number of stable and unstable equilibrium choice models can be constructed in terms of these gradients. This analysis will make use only of the avoidance gradients. That is, the choice theory model employed is of the avoidance-avoidance type. This model is represented in Figure I.⁹ (At present, the reader should ignore the graph labels in

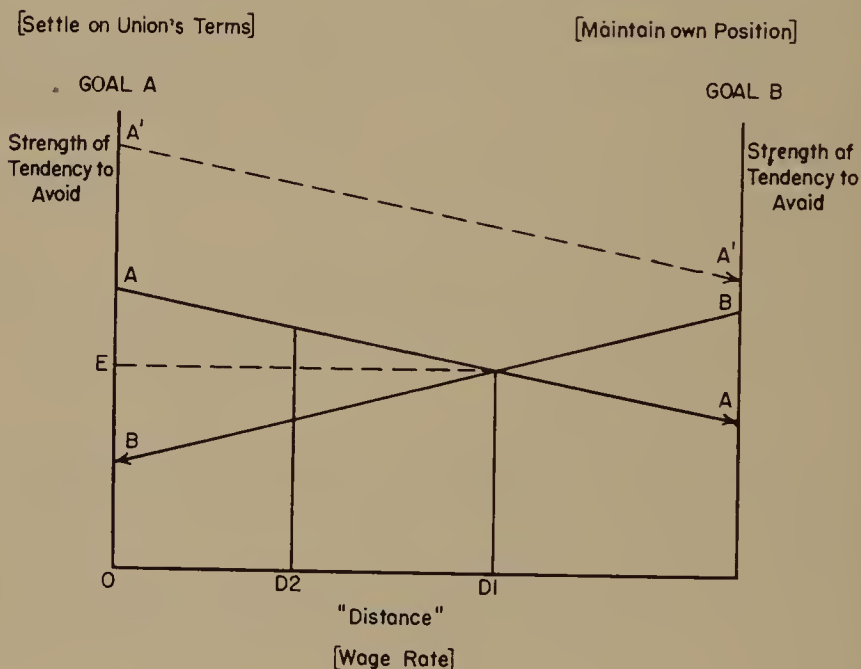


FIGURE I

brackets.) In Figure I, view the goals *A* and *B* as negative goals. The function *AA* is an avoidance gradient for goal *A*, representing the increasing strength of the tendency to avoid goal *A* as the subject gets closer to it. And, analogously, *BB* is the avoidance gradient for goal *B*. At a position such as *D1*, the strength of the tendency to avoid goal *A* is equal to the strength of the tendency to avoid goal *B*. At such a position, the individual is in a stable equilibrium position. If, in this avoidance-avoidance situation, the individual be momen-

8. See Dollard and Miller, *op. cit.*, for a discussion of the inductive and deductive bases for these hypotheses.

9. The linear assumptions incorporated in Figure I are not significant. For this analysis there is no need for restrictions on the shape of these functions other than their general slopes.

tarily displaced from a position such as $D1$, say to $D2$, he will tend to return to $D1$. The reason for this is that in a position such as $D2$, the strength of the tendency to avoid goal A is stronger than the strength of the tendency to avoid goal B . Hence there is a net avoidance tendency operating to drive the individual toward goal B , i.e., back towards position $D1$. An analogous argument applies to displacements on the other side of $D1$, closer to goal B .¹

Suppose the individual to be in equilibrium at position $D1$. Further suppose that, due to a change in environmental circumstances, the individual's avoidance gradient AA shifts upward to occupy the position $A'A'$ (the dotted line function). In this instance, the avoidance gradient to goal A lies everywhere above that to goal B , hence everywhere there is a net tendency to avoid A . In this circumstance, the individual will elect goal B . It is only in such a circumstance that, in spite of the avoidance-avoidance nature of the choice, one goal will be immediately elected.

It will be observed that the model determines two variables: (1) the distance of the subject from each goal, in equilibrium; (2) the strength of the tendency to avoid both goals, which is the same for both goals in equilibrium. In Figure I and for position $D1$ this is measured by OE . Regarding this latter variable, conflict choice situations involving this kind of behavioral equilibrium are apt to be an uncomfortable experience for the individual. Such situations give rise to a drive stimulus commonly termed "tension" (or anxiety, etc.). The level of tension experienced in such an equilibrium choice situation is an increasing function of the strength of tendency to avoid which the individual feels in equilibrium, and of the length of time the subject is in the equilibrium situation. High or long-experienced levels of tension are apt to be associated with a variety of aberrant behaviors, e.g., reduced learning capacity, reduced ability to pay attention, etc. The theory of conflict choice involving this kind of behavioral equilibrium suggests that if there is available some compromise behavior, other than electing either of the two negative goals, the individual should be expected to choose it. (For example, escape from the choice situation altogether, or some compromise response "between" the two goals.) We shall have use for these considerations subsequently.

1. It may be noted with respect to Figure I, that if A and B are viewed as positive goals, and AA and BB as approach gradients, then the model is of the unstable equilibrium type.

IV

Given this general background, let us turn to what is the first task of a theory of negotiation. This is to explain why it is that, in a type of situation such as collective bargaining, the parties negotiate at all.² Negotiation is not after all the only way in which the transaction might be mediated. Why should not the parties elect some alternative method, e.g., "take-it-or-leave-it," for mediating the transaction in question? In seeking an answer to this question it will be helpful to investigate a simple alternative model.

Assume a well organized union whose members comprise the major portion of the company's work force. Further assume that the company would have extreme difficulty in replacing the union members should they strike, and that, in such an event, the union members would experience extreme difficulty in finding alternative employment. Now assume a market procedure as follows: The union unilaterally states its terms (demand for a wage increase) on a take-it-or-leave-it basis, no compromise possible, and with the understanding that should the company elect to leave-it, a strike would be a certainty.³ In this situation, two choices are available to the company:

- (1) insisting upon its own position, the existing terms.
- (2) settling on the union's terms.

Economic choice theory might suggest that the company's choice in this situation should turn simply upon the maximization of expected utility, taking into account such factors as the relative probability and cost of a strike, on the one hand, and the probable cost of paying the rate increase on the other. That is, the company would immediately elect one goal or the other, depending upon the outcome of such a "calculation."

However, this is a conflict choice situation conforming to the avoidance-avoidance model. The company confronts two negative goals, i.e., there is an expectation of cost consequent upon the elec-

2. Such considerations as the existence of a statutory requirement to bargain are not at this point relevant.

3. This was, indeed, very much the way in which "collective bargaining" was carried on in this country in the early nineteenth century. In so far as the point made in this section is concerned, we might alternatively suppose the company to quote the terms, unilaterally. The following analysis applies to both cases.

In a negotiation situation, the union's initial statement of terms would not be its final offer, but rather a "gambit" price enough above its final offer to leave room for bargaining. This, however, is a take-it-or-leave-it situation, and the initial statement of terms is the final one.

For a discussion of the various prices with which a bargainer may be subjectively concerned, see Shackle, *op. cit.*

tion of either. Company choice 1 exposes the company to a strike and the costs associated therewith. Company choice 2 means, *ceteris paribus*, loss of profit due to paying the higher rate.⁴ In consequence, one should expect the company to elect neither of its available choices immediately. Rather, the company will remain uncertain, unable to "make up its mind," i.e., in a sort of behavioral equilibrium between the two. It will also be recalled that in such a choice situation the company should be expected to perform some compromise response *were such available*. Thus the company might prefer that the rules of the market game permitted an additional choice, viz.:

(3) seeking a compromise, via negotiation.⁵

The market "game" as I have here presented it (the-take-it-or-leave-it model) is what I shall term an unnatural, purposive game.⁶ I define a purposive game as one in which the relation between the game and its associated economic transaction is such as to make the game a means to an end, namely, to mediate the economic transaction.⁷

The take-it-or-leave-it game is considered unnatural in the

4. This way of looking at the matter is in line with Hicks's depiction of the employer's position in the bargaining situation, viz.: "... either he must pay higher wages than he would have paid on his own initiative (and this generally means a prolonged reduction in profits) or on the other hand must endure the direct loss which will probably follow from a stoppage of work." See J. R. Hicks, *The Theory of Wages*, pp. 141 et seq. See also Chamberlain's elaboration of the concept of "bargaining power" in Neil W. Chamberlain, *Collective Bargaining* (New York, 1951), chap. 10. Chamberlain adopts a modification of Hicks's view.

5. Or perhaps the company might attempt escape from the choice situation entirely, e.g., obtain a court indictment of the union as a criminal conspiracy, or seek injunctive relief against the strike, etc. From this point of view, the existence of a statutory requirement to bargain is significant, i.e., as blocking off certain potential escape routes.

6. One might, of course, construct alternative models. One feature of the model presented was a certain asymmetry in the amount of information available to the parties. The union made its demand with no knowledge of the company's position whereas the company made its choice with perfect knowledge of the union position which would lead to a strike if rejected. We might remedy this feature by providing that the union and the company are to state their terms simultaneously (e.g., the wage rate) on a take-it-or-leave-it basis. This would now be a two person, two strategy game in which the union had available strategies analogous to those available to the company in the example cited. Provision might also be made that the parties must reveal information with respect to their choices simultaneously. Investigation of such additional models does not, however, seem to lead to conclusions substantially different from those involved in the simple model above discussed.

7. This is in contrast to certain parlor games in which the relationship between the game and its associated economic transaction (wager) is the reverse of this. In these games, the game itself is the end, and the economic transaction a means to the end, i.e., to lend interest to the playing of the game.

The distinction between purposive and nonpurposive games is an important one for the reason that the two classes may differ with respect to properties crucial

avoidance-avoidance choice situation because, in this situation, the player is highly motivated (by tensions generated in the choice situation) to seek strategies in addition to those made available by the rules. Of course, in any game, increasing the number of opportunities available may, *ceteris paribus*, increase the value of the payoff. However, I do *not* here argue that, in the case of an unnatural game, the player seeks an additional strategy (negotiation) because he may thus probably be assured a more favorable outcome of the game.⁸ Rather, I argue that the avoidance-avoidance conflict choice situation is inherently and generically of such a nature that the game take-it-or-leave-it must give rise to strong motivations to discover alternative responses. In this situation the individual cannot immediately make up his mind which goal to elect. He is in a kind of behavioral equilibrium such that strategies other than those available in take-it-or-leave-it are psychologically necessary if the game is to be an appropriate (natural) one for mediating the transaction in question.

I conclude, then, that in the kind of bilateral monopoly situation represented by collective bargaining, the parties would probably prefer to play the game negotiation rather than the game take-it-or-leave-it. This conclusion would seem to be somewhat at odds with Schelling's approach to the theory of bargaining. Schelling's major point is that an important class of tactics in bargaining is that class which enables a player to convert the choice confronted by his adversary into one of the take-it-or-leave-it variety.⁹ However, in the light of the above analysis of why the parties negotiate, this tactic should not be of prime importance (at least at the outset) in many negotiations. I have argued that (for rational or other reasons) the parties choose to play the game negotiation rather than the game take-it-or-leave-it in certain kinds of choice situation (i.e., conflict choice of the avoidance-avoidance type). It would not seem plaus-

to analysis. Thus, for example, one would hazard the guess that purposive games are more apt to give rise to conflict choice situations involving the experience of uncomfortable tensions than are parlor games. Consequently, it would seem more important to consider this possibility in an analysis of purposive games. On the other hand, a game theory elaborated chiefly with respect to nonpurposive games might neglect this possibility.

8. Such a rationality principle argument can be made. Suppose that a solution "no deal" has an extremely low value. Suppose further that, as contrasted with negotiation, take-it-or-leave-it greatly increases the chance of such an outcome. Then one might choose to mediate the transaction via negotiation precisely for the reason that this increases the probable payoff.

9. That is, player *A* contrives to so bind himself that *B* is convinced that *A* will not concede, i.e., *B* is confronted with a take-it-or-leave-it offer. This tactic may well be of great importance in some kinds of bargaining situation. (See Schelling, *op. cit.*)

ible, then, that a primary negotiation tactic should be an attempt immediately to convert the game negotiation into one of the take-it-or-leave-it variety.

V

The next task is an analysis of the negotiation process itself. For this purpose let us again utilize the avoidance-avoidance paradigm. Assume that the sole bargaining issue is the wage rate. We shall explicitly consider just one side of the situation, i.e., the choice problem as confronted by the employer and the negotiation task as seen by the union. The principles involved will likewise apply to the situation seen from the other side. Figure I now represents the employer's position, the goal *A* being "settle on union's terms," and the goal *B* being "maintain own position." Since this is a negotiation situation, the union's demand and the company's position must be viewed as "gambit" prices, i.e., announced prices, leaving room for negotiation, rather than necessarily final prices. We need an appropriate conceptualization of the concept "distance" (from goal) — the horizontal axis in Figure I. In this context, distance will have to be interpreted in some subjective sense rather than in a literal spatial sense. Generally speaking, postulation of this kind of behavioral equilibrium results in dividing choice behavior into two stages: (1) the initial stage of deliberation or "making up the mind"; (2) a subsequent stage of actual overt choice behavior which will take place once stage (1) is resolved.¹ The concept of "distance" from goal *A* shall be interpreted as "nearness to having made up one's mind" to choose goal *A*. This concept of distance is, perhaps, susceptible to operational definition only in terms of an individual's responses upon interrogation. It seems probable that in many instances a subject would be able to testify on this matter only in an ordinal sense. In the collective bargaining situation, the two goals represent certain wage rates (the announced positions of the parties), and hence the distance between the two may be thought of as scaled in terms of such rates. Suppose the company's announced position to be \$1.50 and the union's announced position to be \$2.00. An equilibrium position halfway between would be represented by a rate of \$1.75. The individual in such an equilibrium position will be supposed ready to

1. As Macfie has noted, such a division of the choice problem into two stages, which is not characteristic of choice theory in economics, is characteristic of much psychological choice theory. See A. L. Macfie, "Choice in Psychology and as Economic Assumption," *Economic Journal*, LXIII (June 1953), 352-67.

accept a rate of \$1.75, *if there were any way to do so*. The "distance" between the equilibrium position and the goal \$2.00 is \$.25 — and the individual may be thought of as this near to having made up his mind to elect the goal \$2.00.

Again with reference to Figure I, the goals confronting the company are negative goals. To elect the goal *B* may result in a strike. The expected cost calculation on this score must presumably be compounded of two parts:

(1) an estimate of the probable cost to the company of a strike, should it occur;

(2) an estimate of the probability that the company's insisting upon its own position will indeed eventuate in a strike.

Goal *A* is also a negative goal, the cost being lower profits in consequence of higher wage rates, *ceteris paribus*. The company may be viewed as initially in equilibrium at position *D1* — unable to make up its mind either to insist upon own position or settle on the union's terms.

The union will accomplish its negotiation task by means of information exchanged during the course of negotiations. One important aspect of this task is to move the company's equilibrium position in a direction favorable to the union, that is, toward the company's goal *A*, i.e., settling on the union's terms (see Figure I).² Two classes of negotiation tactics which will have this effect must be distinguished:

Class I: Tactics which will raise the company's avoidance gradient to the company's goal *B*, i.e., insisting upon its own position.

Class II: Tactics which will lower the company's avoidance gradient to the company's goal *A*, i.e., settling on the union's terms.

Subsequently, tactics of Class I will be referred to simply as raising the avoidance gradient, and tactics of Class II will be referred to as lowering the avoidance gradient, the goals of reference being understood.³

2. The other important part of the task is for the union to communicate information regarding its own equilibrium position without at the same time destroying whatever negotiation power it may have. This matter is discussed subsequently.

3. Adaptation of the avoidance-avoidance model to analysis of collective bargaining results in a conceptualization of that process similar in many ways to approaches suggested in the literature. See, for example, Hicks, *op. cit.* Compare also Chamberlain's definition of "bargaining power." (See n. 4, p. 78; see also his *Collective Bargaining*, chap. 10.) Using Hicks's analysis as a point of departure, Chamberlain defines *A*'s bargaining power vis-à-vis *B* as the ratio of the cost to *B* of disagreeing with *A* on *A*'s terms to the cost to *B* of agreeing

It might be objected that these classes of tactics are not really independent. That is, that a tactic which will increase the company's tendency to avoid its own position (Class I) will, at the same time, decrease the company's tendency to avoid the union's position (Class II). The implications of this way of putting the matter are misleading. It is true, of course, that a union tactic which will move the company's equilibrium position farther from the company's goal *B* (maintaining its own position) will, at the same time, move that equilibrium position closer to the company's goal *A* (settling on the union's terms). However, tactics of Classes I and II are not distinguished in terms of movements of the equilibrium position, but rather in terms of movements of the avoidance gradients, i.e., tendencies to avoid viewed in the *schedule* sense. In these latter terms, the independence of the two classes seems plausible. The company's avoidance tendencies are based upon calculation of expected cost associated with paying a higher wage rate, on the one hand, and a strike on the other hand. Let us suppose, for example, that the information exchanged during negotiation will revise downward the company's estimate of the cost associated with paying a higher rate (a successful union Class II tactic). There is no *prima facie* reason to suppose that this tactic must necessarily have a Class I effect, e.g., cause a revision of the company's estimate of the probable cost associated with a strike.

VI

An analysis of the negotiation process in terms of this schema is in large part an analysis of the shift parameters relevant to the avoidance functions. Thus, e.g., Party One's various negotiation tactics will be seen as attempts to shift Party Two's avoidance functions so as to move Two's equilibrium position closer to (and if possible coincident with) One's demand.

It has been my intention to present a framework in terms of which negotiation processes may be analyzed. It is not my intention

with *A* on *A*'s terms. The significant thing about Chamberlain's concept of bargaining power is not that it results in a coefficient of that power in any particular situation. Rather, the significant thing is the theory of the negotiation process implied by it. His scheme, unlike much discussion of bargaining, recognizes both what we have termed Class I and Class II tactics. Beyond such similarities to the avoidance-avoidance model, there are also important differences. For example, Chamberlain does not develop any explicit equilibrium theory of conflict choice situations. Thus his scheme does not explain *why* it is that if, in his terms, *A*'s coefficient of bargaining power vis-à-vis *B* is greater than one, *B* doesn't simply settle on *A*'s terms — rather than go on bargaining (as, he indicates, may indeed be the case).

to undertake an extensive analysis of these processes. However, an example to illustrate at least the application of this framework would seem desirable in order to make the nature of such application more concrete.

Consider the problem of the function of the so-called "basic criteria" used in wage negotiations, e.g., ability to pay, productivity, comparative rates, etc. This is an interesting example because of the widespread use of such principles in the course of negotiations, and because of the nature of the discussion which has surrounded the practice.

In many wage negotiations, the participants spend considerable time and energy in gathering data relevant to, and in debating the applicability of, various of these basic criteria. This circumstance would seem to constitute at least *prima facie* evidence that these principles must play some functionally important role in the wage negotiation process. However, much discussion of the use of these principles takes a position, in one way or another, that would lead to the conclusion that these principles play no substantively important role in the negotiations.⁴ (For example, a substantively important role in the sense of serving to define the "real" wage objectives of the parties, etc.)

Looked at from the point of view of the negotiation theory herein developed, it becomes apparent that use of such principles may well serve important tactical purposes. Let us consider the comparative rates criterion. Suppose the union to contend that its own demands are justified and, indeed, minimal in the light of rates prevailing in a somehow defined comparable labor market area. What function might this communication serve?

It might actually serve to change the company's mind with respect to this aspect of its environment, i.e., serve to convince the

4. For example, Lindblom, viewing this matter from the point of view of union wage aims, deems these principles mere "window dressing," utilized by the union for "expediency's sake." (Cf. Charles E. Lindblom, *Unions and Capitalism*, New Haven, 1949, chap. III.) Bloom and Northrup explain their lack of attention to these wage criteria on the ground that "... these criteria are basically rationalizations pressed into service to support demands which need justification," and point out that these "... slogans should not, however, be confused with the practical realities of union-management wage determination. . . ." (Cf. Gordon F. Bloom and Herbert R. Northrup, *Economics of Labor and Industrial Relations*, Philadelphia, 1950, p. 204.) Reynolds likewise views these criteria as "rationalizations," and points to the "lack of substance" and "frailty" of arguments involving these criteria. (Cf. Lloyd G. Reynolds, *Labor Economics and Labor Relations*, New York, 1954, pp. 575, et seq.) This list of citations could be extended, but the sample is representative of the point of view here in question, and sufficient to make the point.

company that its rates were indeed low on a comparative basis. If this were the effect, and if the labor market area selected for the comparison contained the company's important competitors, then the comparative rates argument might serve as a Class II negotiation tactic (decreasing the company's tendency to avoid accepting the union position). This would be so to the extent that the company had previously avoided accepting the union position for fear of being thereby put at a competitive disadvantage.

However, there is the additional and perhaps more likely possibility that this comparative-rates argument has its major effect as a Class I negotiation tactic. From this tactical point of view, the important functional role of the union's comparative-rates argument is less to communicate information to the company about the company's wage-rate environment than to communicate information to the company about the *union's beliefs* on this matter.⁵ That is, the argument may serve to convince the company *that the union believes itself (the union) to be correct in asserting its demands to be minimal in accordance with the comparative rates criterion*. If this were the effect of the argument, then it would constitute a Class I argument. The reasons why this is so are as follows:

(1) It will be recalled that the expected cost calculation tending to make the company avoid insisting upon its own terms was comprised of two elements: (a) the estimated cost of a strike, should it occur; (b) the probability that a strike would indeed eventuate.

(2) The company may believe that, to the union officials, because of their status as elected officials who must satisfy the membership, comparative rates parity is an extremely important matter.

(3) If the comparative rates argument had the effect hypothesized (see italicized passage), it would be apt, in the light of (2), to revise upward the company's estimate of the probability that a strike would eventuate should the company insist upon its own position. And this upward revision would shift the avoidance gradient upward, moving the company towards the union's demand.⁶

Other wage "principles" and, indeed, any other negotiation tactic (e.g., striking, taking a strike vote, appealing to the public, etc.) are susceptible to analysis in terms of the conceptual scheme here pre-

5. This does not carry the implication that the union is aware of or intends this effect.

6. Note that the factor of the probability that a strike would occur, not the factor of the cost of a strike should it occur, has been the expected cost element influenced by this particular tactic. In general, one should expect that negotiation tactics might operate upon either or both elements.

sented. I shall not here attempt a demonstration of this assertion.⁷

VII

To complete this conceptual framework for the analysis of negotiation, it is necessary to explain why and in what ways the negotiations may come to an end. There appear to be at least five analytically distinguishable ways in which the negotiations may end, viz.:

- (1) Compromise
 - (a) Unilateral
 - (b) Bilateral
- (2) Noncompromise
 - (a) Type I
 - (b) Type II
- (3) Breakdown

Although we have thus far been looking at the negotiation process largely from one side of the table, i.e., possible effects of the union's

7. The above analysis suggests that negotiation arguments, e.g., those concerning the basic criteria, may take on a new significance if viewed from what might be termed a "tactical" rather than a "substantive" point of view. (Regarding this distinction, see my "Regarding the Determinants of Union Wage Policy," *Review of Economics and Statistics*, XXXV, Aug. 1953). This suggestion is in line with that made by Schelling, *op. cit.* Indeed, the interpretation of the comparative rates argument as a Class I negotiation tactic makes it similar in effect to that particular class of tactics to which Schelling intended to draw attention. There are these important differences in emphasis, however: (1) The union does not, in my example, bind itself in an all-or-none fashion by converting the game of negotiation into one of take-it-or-leave-it. In my example, the effect is a much more subtle one, tending to increase (in the company's eyes) the degree of adherence of the union to its position. (2) The effect of the tactic is not to result either in a unilateral compromise (by the company to the union position) or no deal. Rather, it serves just to move the company's equilibrium position somewhat closer to the union's terms. Thus it may help to set up the conditions necessary for a compromise. A good bit of argument remains before we can explain how compromise is possible in a purposive game such as collective bargaining over terms and conditions of employment.

Martin Fishbein has suggested another interpretation of the role of the basic criteria in wage negotiations. (See his *A Social-Psychological Approach to Collective Bargaining*, unpublished B.A. thesis, Reed College, 1957). Fishbein draws attention to Nicholas Pastore's suggested modification of the Dollard (*et al.*) frustration-aggression hypothesis. According to Pastore, an aggressive response to frustration is less likely if the subject perceives the obstacle to his goal achievement as "reasonable" (i.e., nonarbitrary). Fishbein suggests that through widespread use in the process of collective bargaining, the basic criteria may be becoming institutionalized as "reasonable" obstacles to union-management goal achievement (e.g., lower wage rates, in the case of the company goal), thereby decreasing the probability of aggressive responses in the bargaining context.

negotiation tactics on the company's equilibrium position, it will be recalled that the analysis is intended to apply symmetrically to both sides of the table. From now on it will often be helpful to think in terms of two figures such as Figure I, one representing the union's conflict choice situation and equilibrium position and the other representing the company's conflict choice situation and equilibrium position. In these terms, the status of any particular collective bargaining negotiations at any particular time may be represented by values, e.g., as assembled (for subsequent use) in Exhibit I:

EXHIBIT I¹

	Status 1	Status 2	Status 3	Status 4
Union's Announced Demand	\$1.25	\$1.25	\$1.25	\$1.25
Union's Equilibrium Position	1.15	1.12	1.20	1.12
Company's Equilibrium Position	1.10	1.12	1.13	1.05
Company's Announced Offer	1.00	1.00	1.00	1.00

1. All figures represent hourly rates.

By the outcome "compromise" I mean simply termination of the negotiations with an agreed price, i.e., with a position to which both parties assent. In outcome (1) (a), unilateral compromise, the agreed price is the same as the announced position of either party.⁸ Unilateral compromise will be the outcome if the negotiation tactics of one party, say the union, succeed in shifting the equilibrium position of the other party to its (the union's) announced position.

Compromise, unilateral or bilateral (to be discussed subsequently), may well follow the occurrence of a strike, as well as numerous other negotiation tactics. It is important in the analysis of negotiation to realize that what I term the "legitimate" strike is not an instance of the "breakdown" of negotiations (category 3 in the above classification of terminations). The legitimate strike, as a deliberate attempt on the part of the union in this way to increase the company's tendency to avoid adherence to its own position, is perhaps best viewed as an effective way of communicating information. Previously the company has experienced the threat of a strike. The actual occurrence of the strike raises to a value of one the company's estimate of the probability that a strike will occur if the company insists upon its own position. Hence, as an integral part of the negotiation process (not as a result of the breakdown of that process), a legitimate strike may be instituted by the union if the union thinks that the company:

(a) underestimates either the cost to it, or the probability of a strike, and

8. This position need not be the initial gambit price.

(b) would probably assent to the union's demand were a correct estimate made.⁹

If, in such a case, the union were correct on points (a) and (b), the legitimate strike may be followed by a compromise, the company modifying its position.

By definition, in outcome (1) (b), bilateral compromise, the agreed price will lie somewhere between the announced positions of the parties.

The development of a good theory of bilateral compromise is at once one of the most important and most difficult problems in the development of a general theory of negotiation. It will not do simply to assume that bilateral compromise is in some sense a "natural" outcome of the negotiation process. A negotiation theory must explain how it is possible for the parties to achieve bilateral compromise.¹

The conflict choice analysis indicates that it is a necessary condition for the occurrence of bilateral compromise that the equilibrium positions of the parties be brought, via negotiation, to the same position. For example, with reference to Exhibit I, Status 1 represents a situation in which bilateral compromise is not possible, and Status 2 one in which it is possible. One definite task to be accomplished by the exchange of information in negotiation is to bring this equality (necessary but not sufficient for bilateral compromise) about. Further, the parties must somehow be informed of the similarity of their equilibrium positions if the fact of this similarity is to create a compromise route. But how is this information to be conveyed? Suppose with respect to a particular negotiation that, with reference to Exhibit I, Status 2 obtains. That is, at the time in question, the conditions necessary for bilateral compromise obtain. Suppose now the union simply to announce its equilibrium position, i.e., \$1.12 (which is the same as the company's position, but the union does not know this). This announcement, rather than resulting in immediate bilateral compromise, may serve to stiffen the company's position by causing the company to revise downward its estimate of the probab-

9. This would suggest that *A* might protect himself against a legitimate strike by *B* by convincing *B* that he, *A*, held the highest possible estimate of both the probability of a *B* strike and the cost of a *B* strike when it occurred. From this point of view, perhaps the present tendency in international relations to lay great stress upon the likelihood of atomic war and the tremendous destruction sure to follow serves a useful purpose.

1. There is also the question of *why* the parties seek bilateral compromise. The discussion in Section IV of this paper may be taken as directly relevant to this question.

ity that adhering to its own announced position will result in a strike.² In terms of Exhibit I, for example, the union's announcement of its position may cause the negotiations to shift from Status 2 (Status at time of announcement) to Status 4 (where the necessary conditions for bilateral compromise are no longer met).

In the light of these considerations, what is the second major function of the exchange of information during negotiations becomes clear. This second function is to enable *A*, for example, to inform *B* of his (*A*'s) equilibrium position in such a way as not to destroy *A*'s bargaining power by the very process of the communication.

The question of precisely how such a communication may be accomplished is an important one, and one to which I cannot deliver a definite answer. However, it should be noted that M. W. Reder has made some suggestions of interest from this point of view.³ On Reder's view, the problem in the bargaining duel is to define "fair treatment." This is so because, although within limits any settlement is better than no settlement, the parties "... do not wish to appear over-eager for agreement (lest they tempt the other party to take advantage of them), and they wish to be treated "fairly."⁴ Reder then points out that among the principal ingredients of equitable treatment will be Ross's notion of orbits of coercive comparison (the comparative rates criterion). The other basic criteria may likewise be viewed as relevant from this point of view. This suggestion may be interpreted somewhat as follows. Suppose the union's announced price to be \$1.25 and its equilibrium price \$1.12. As previously indicated, the union's simple and direct communication of information regarding its equilibrium price may be destructive of its bargaining power. However, the union may be able to argue in terms of the comparative rate (and other basic wage) criteria in such a way as to convey information that some price less than its announced demand would still be considered acceptable as "fair." What is here implied,

2. This point has been made elsewhere in the literature, for example by Chamberlain (see his *Collective Bargaining*, *op. cit.*, chap. 10). In his terms, *A*'s attempt to "increase bargaining power" by reducing the magnitude of a demand (which will reduce *B*'s cost of agreement) may fail to increase bargaining power because it may also decrease *B*'s notion of the cost of disagreeing with *A*. In general Chamberlain views alterations in the magnitude of demands, as well as all other bargaining and negotiation tactics which, in our terms, would be viewed as shift parameters in the avoidance functions, as ways in which to change the magnitude of bargaining power. It would seem better to view these tactics as the utilization of bargaining or negotiation power, whatever its magnitude, to achieve an objective.

3. See his "The Theory of Union Wage Policy," *Review of Economics and Statistics*, XXXIV (Dec. 1952), 34-45.

4. *Ibid.*, p. 39.

then, is that an "orderly" retreat (based on "principle"), from an asking price to a "fair" price, may not be destructive of bargaining power (i.e., on Reder's terms, will not cause the other party to "take advantage" of the party so involved).

Turning now to the outcome noncompromise, let us mean by this the termination of negotiations without an agreed price. Such a termination of negotiations implies some sort of strike or lockout (before or after such termination), but not a permanent severance of the relationship between the parties.⁵ Suppose that, after negotiations have terminated and the union is out on strike, the strike is broken. That is, there is a nonunion-ordered return of union (and nonunion) members to work on company terms to which the union has not assented. In such a situation, the union may ultimately sign a contract on company terms, if only to salvage its status as bargaining agent. In terming such an outcome noncompromise, i.e., one in which there is no agreed price, I mean, in effect, that it is an outcome which is the result of a breakdown of an "alliance" on one side — in this case, the union and its members. Such an outcome (although it involves in some sense an "agreed" price), does not represent a compromise in our sense, i.e., a compromise between two parties which retain their organizational integrity. Rather, it represents the destruction of the organizational integrity of one of the parties.

Within the general outcome category noncompromise, it is helpful to distinguish:

- (2) (a) Type I,
- (2) (b) Type II.

This distinction is based upon the two general functions which the exchange of information in negotiations is supposed to serve. The first of these, it will be recalled, was to alter the equilibrium positions of the parties by shifting the avoidance gradients up and down. It is clear that negotiation may fail to result in compromise because this function has not been served in such a way as to set up the conditions necessary for compromise — i.e., to bring the equilibrium positions of the parties into consonance. This is a type I noncompromise outcome.

Even if the necessary conditions for compromise have been achieved, however, a noncompromise outcome may still be the result if the parties are unable mutually to inform each other of this fact so as to turn the fact of this equivalence into an actual compromise

5. The outcome noncompromise may follow a legitimate strike, as well as other negotiation tactics. This will be the case if, in the light of the information conveyed by the legitimate strike, the company still does not agree to the union position.

route. A failure on these grounds constitutes a type II noncompromise outcome.

It would appear that the institution of mediation is primarily useful in preventing negotiation failures of the type II noncompromise variety. The parties while not free, for reasons explained, to communicate information directly regarding their true equilibrium positions to each other, are free to do so to a neutral third party, the mediator. The latter is in a position with this knowledge to inform the parties if and when they have indeed achieved the condition necessary for bilateral compromise. Thus the mediator, in a sense, performs the second of the two major functions of the exchange of information in the negotiation process.⁶

Let us turn finally to the possibility of breakdown as an outcome of the negotiation process. It will be recalled that the avoidance-avoidance model determined not only the individual's equilibrium distance from the negative goals, but also the strength of the individual's tendency to avoid these goals, which, in equilibrium, is the same for both goals. (Referring again to Figure I, for distance $D1$ the strength of tendency to avoid is measured by OE .)

The possibility of the outcome breakdown is suggested by the second of these two variables. For, as pointed out in Section III, the conflict choice situation is apt to give rise to tension or anxiety. Further, the amount of tension experienced in this choice situation is apt to be an increasing function of the strength of tendency-to-avoid experienced in equilibrium. High levels of tension may lead to various forms of aberrant behavior, e.g., reduced learning capacity, reduced ability to pay attention, etc. Such behavior might, *per se*, reduce the chances of finding a compromise solution via negotiation. Beyond this, a part of such behavior might be simply to "bolt" the bargaining situation in a precipitate and nondeliberate fashion.⁷ Such a termination of the negotiations is what we here term a breakdown of the negotiations — to be distinguished from the outcome noncompromise.

6. This would suggest that perhaps mediation (of a sort) should be a part of many negotiation proceedings *from their inception*. Thus it might be required that the parties continuously submit, during the course of negotiations, information regarding their true equilibrium positions to a neutral third party. It would be understood and agreed that the sole function of this neutral third party would be to receive the information in question and announce the fact of conditions necessary for bilateral compromise — if and when such conditions should eventuate. The value of employing mediation of this sort from the outset is that it might prevent the parties from destroying a condition necessary for bilateral compromise in their unaided attempts to discover the fact of its existence.

7. Unlike, for example, a calculated legitimate strike.

It should be emphasized that the argument here is not that, typically, collective bargaining negotiations will involve the kind of emotional content probably leading to breakdown.⁸ Rather, the position is this: If, indeed, the avoidance-avoidance paradigm is conceptually appropriate for analysis of the negotiation choice problem, then the possibility of breakdown must be taken into account in the analysis.

The above considerations suggest the possibility of an important asymmetry in the effects of the two classes of negotiation tactics. It will be recalled that Class II tactics, as used by, say, Party One serve to decrease Party Two's tendency to avoid agreement with One. Such an effect will drive Two closer to One's position. It will also serve to *decrease* the amount of tension experienced by Two in equilibrium. Class I tactics, on the other hand, serve to increase Two's tendency to avoid his own position. Such an effect will likewise serve to drive Two closer to One's position, and in this sense may be equally efficacious with the Class II tactic. Unlike the Class II tactic, however, the Class I tactic will increase the degree of tension experienced by Two in equilibrium. On the basis of the earlier argument then, negotiations featuring heavy emphasis upon Class I tactics should, *ceteris paribus*, be more apt to terminate in breakdown than negotiations featuring relatively heavier emphasis upon Class II tactics.

VIII

One final point warrants brief attention. In thinking about general conceptual schema adequate to the analysis of negotiation, should not some place be given to game theory? Attempts have been made to analyze aspects of bargaining in these terms.⁹ However, in this writer's opinion, the game theory format is essentially inappropriate to the analysis of negotiation. This is for reasons beyond the often mentioned nonzero sum nature of most negotiated games. Game theory emphasizes a rationality-type solution with the calculation of optimal strategy elaborated with respect to a supposedly known or

8. Although this factor is probably important in some negotiations. Edward Peters (*Conciliation in Action*, New London, 1952, pp. 28-29), after warning against an overemphasis upon the emotional factor and pointing out that disputes are based first and foremost upon the relative strengths of the parties, goes on to point out: "This does not mean that deadlocks are not marked by a tense emotional atmosphere. It does mean taking into account that the contestants usually exhibit strong emotions because of the issues between them; and not, as is often supposed, that there are issues between them because of angry passions and deep emotions. The emotions are not the cause of the deadlock. They are one of its effects." Agreeing with this general position, one may then go on to argue that these emotions themselves may have additional effects.

9. See, for example, Wagner, *op. cit.*

somewhat arbitrarily assumed payoff matrix. But in most negotiated, purposive games, precisely the major task of the exchange of information during negotiation is to change the negotiators' perception of the values comprising the payoff matrix. Herein lies the essence of the analysis. Elaboration of techniques for the calculation of optimal strategies on the basis of known payoffs would not seem to add much to the analysis of this type of situation.

CARL M. STEVENS.

REED COLLEGE

MONETARY SURVEYS AND MONETARY ANALYSIS*

By CLAUDIO SEGRÉ

I. Introduction, 98. — II. The theoretical background of monetary surveys, 99. — III. Money supply analysis and monetary financing in the open economy, 102. — IV. Inflationary and deflationary impulses and reactions, 105. — V. The problem of sectoring, 108. — VI. A simple model of monetary survey, 109. — VII. Reserve money analysis: the model, 111. — VIII. Reserve money analysis: interpretation of the accounts, 113.

I. INTRODUCTION

Until not long ago economists seemed to have discarded the “monetary approach” to inflationary and deflationary phenomena as an obsolete analytical tool. Reliance on the “income-expenditure” approach had become widespread. As a consequence, between the development of monetary theory and the practical application even of established theoretical propositions a gap had appeared. The gap now seems to have been bridged at last: monetary analysis, however, has had to amend its ways, and no longer concentrate on such simple tools as the ratio of reserves to sight liabilities of the central bank, gold flows, and changes in the money supply. Current efforts are instead directed at analyzing the whole network of financial transactions so as to show the interrelations between the various sectors’ operations and, for each sector, between sources and uses of funds.

The most complex type of flow-of-funds accounts, developed by the Federal Reserve,¹ offers an analysis of all intersector transactions in money terms relating not only to current production, but also to the exchange of financial claims and of existing assets. Such flow-of-funds accounts provide an altogether new alternative to conventional national income accounts, as they focus on the determinants, composition and effects of expenditure decisions by the various sectors, rather than on the productive performance of the economy.

In other countries, however, a similar shift in emphasis has been obtained without so extensive an elaboration of statistical material. These other presentations try simply to integrate a somewhat modi-

* The author has benefited greatly from discussion with Professor Robert Triffin, to whom he also owes thanks for constant encouragement. Professor William Fellner read an earlier draft and offered most helpful suggestions.

1. Board of Governors of the Federal Reserve System, *The Flow of Funds 1939-1953* (1955).

fied version of the national income accounts with a record of the changes in indebtedness and liquidity resulting from the financial transactions of the single sectors. In spite of their less ambitious scope — and perhaps because of it — their usefulness in analyzing short-term tendencies in the economy seems to have been at least equal to that of the Federal Reserve accounts.

Because the operations of the monetary and banking system are so important in intersector financing their central position in such surveys is justified. But they provide no more than a nucleus which leaves wide scope for variety of presentation.² The simpler types of surveys, such as those adopted in the Latin American countries, limit themselves to distinguishing money creation “of internal origin” from money creation due to balance-of-payments developments. Others engage in more complete presentations, often extending to nonmonetary items of financing derived from the accounts of financial institutions other than banks, from the balance of payments, from the government accounts and from a record of transactions on the capital market.

Co-ordination with national income statistics has been achieved by the Central Planning Bureau of the Netherlands, by the Bank Deutscher Länder and the Institut für Wirtschaftsforschung in Germany and by the Ministry of Finance in France. The differences between these monetary surveys consist mainly in the greater or lesser detail in which *sources and uses of funds* (the rows in the matrix of intersector financing) and *sectors* of the economy (the columns) are shown. Such differences are justified not only by the varying availability of data, but also by the relevance of particular institutional distinctions among the components of the monetary and financial mechanism for analyzing the short-term movements of the economy.

II. THE THEORETICAL BACKGROUND OF MONETARY SURVEYS

Monetary surveys are, then, expanded presentations of the national accounts; the main difference from the conventional national income models is the stress laid on a particular question, namely how the income-expenditure disequilibria of the various sectors are

2. A review of currently published surveys has been presented by the Statistics Division of the International Monetary Fund; fifty-seven surveys are analyzed from the viewpoint of sources and methods, and bibliographical references are given for each. See: Hicks, Dorrance and Aubanel, “Monetary Analyses,” *IMF Staff Papers*, Feb. 1957. The Fund’s paper provides not only invaluable basic material but also a conceptual framework which will remain the primary point of reference for future studies on this topic.

financed. But the interest of the system is not in the quantitative description of such disequilibria and of their financing; it is rather in the presentation of the dynamic interrelationships between the changes in the various items.

Monetary surveys do not provide necessarily direct information about the liquidity and the indebtedness of various sectors, which are stock concepts; however, they indicate clearly the direction in which the latter are changing through time. It is important to note that this type of analysis is by no means linked to assumptions of the "naive quantity theory of money" type: on the contrary, in the income-expenditure approach adopted here, expenditure decisions are considered to be the prime determinants of change. Yet it is recognized that liquidity and indebtedness may and will affect such decisions as a constraint on investment or as an incentive to risk-taking or in still other ways which have become explicit in recent writings on monetary theory.³ The monetary element is therefore prominent in this approach, but the link with expenditure decisions is flexible and indirect, rather than rigid and direct.

The framework is also particularly fit for analyses on Robertsonian and Swedish lines. The concept of divergence of *ex-ante* magnitudes and their reconciliation *ex-post* is a tool widely used in theoretical monetary analysis; can this process, however, be traced through the observation of monetary and banking statistics, obviously reflecting *ex-post* magnitudes only? To a certain extent we can actually isolate developments within the monetary system that provide significant evidence on the adjustment of *ex-post* magnitudes in successive periods; from these, in many cases, we can then draw with the help of circumstantial evidence, inferences on the *ex-ante* divergences that have been responsible for the adjustment. Yet it is still a moot point how comprehensive the analysis needs to be for this procedure to be valid. Dr. Holtrop, President of the Nederlandsche Bank, advocates a form of monetary analysis limited in scope to transactions affecting the liquidity position of the various sectors (transactions with money-creating institutions, issue and retirement of short-term Treasury bills, and accumulation and disposal of liquid balances). The Nederlandsche Bank argues that it is actually correct to attribute the responsibility for inflation to the sectors that borrowed from money-creating institutions, dishoarded, or issued second-

3. See, in particular, on liquidity and risk-taking: W. J. Fellner, *Monetary Policies and Full Employment* (Berkeley, 1947), pp. 158 and ff., and on liquidity and consumer behavior: J. Tobin, "Asset Holdings and Spending Decisions," *American Economic Review*, XLII (May 1952), 109-23.

ary liquidities, and its system of analysis is based on this premise. Dr. Holtrop's writings also establish a theoretical case for this practice.⁴ However, one cannot help feeling that in so doing the chain of reasoning is stopped arbitrarily. It may well be that for some situations one can immediately determine the origin and causes of monetary disturbances by looking no further than the transactions affecting the monetary system. Yet this can hold only as long as we assume, for instance, that no important changes occur in the market for long-term capital. It is easy to think, instead, of cases in which the monetary disturbance arises from a changed attitude of investors towards the securities issued by a certain sector: suffice it to mention the effects of tax privileges granted to government securities or any other form of discrimination against private securities. If a cutback in expenditure is to be avoided, the borrowing sector will have to turn to money creating institutions for monetary financing; in simple liquidity analysis such a sector would appear as the origin of the inflationary pressures. It is true that from the point of view of the central bank's responsibility to safeguard the purchasing power of the currency, this is as much as it is necessary to know in order to apply remedial measures. It could hardly be maintained, however, that such action would be satisfactory for the stability of the economy, which could be better protected by acting on the disturbing factor in the capital market.

The problem could be summarized in even more concise terms: the Nederlandsche Bank stresses at all times that "if one analyses monetary phenomena with the purpose of getting some guidance for monetary policy, one must necessarily use a model in which monetary policy can find its place."⁵ We should like to add that not only must monetary policy find its place in the model but also other policies affecting, for instance, the capital market should be introduced into the model. Otherwise we run the risk of misinterpreting as purely monetary a disturbance which should actually be traced further up in the chain of intersector finance. Only then will the monetary survey be able to perform not only its basic function — ascertaining the origin of inflationary or deflationary pressures — but also its second — analyzing the roles played by the various parts of the

4. See particularly the presentations of the Holtrop model given in H. C. Bos, *A. Discussion on Methods of Monetary Analysis and Norms for Monetary Policy*, Netherlands Economic Institute (Rotterdam, 1956); F. J. De Jongh, "Méthodes statistiques d'analyse de la situation monétaire," *Bulletin de la Banque de Belgique*, Mars-Avril, 1956; M. W. Holtrop, "The Method of Monetary Analysis used by De Nederlandsche Bank," *IMF Staff Papers*, Feb. 1957.

5. Holtrop, *op. cit.*, p. 306.

monetary and financial mechanism in favoring or restraining the original disturbance.

The approach followed in this paper will be to discuss and account for the essential features of monetary surveys by examining problems of interpretation in practical monetary analysis. Our attention will be focused first on the shortcomings of money supply analysis in the case of an open economy. The problem of determining the origin and meaning of financial disequilibria will subsequently provide the background for a discussion of the necessary analysis by sources and uses of funds and by sectors; a model based on the integration of national income and financial statistics will then be introduced.

Finally the need for supplementing monetary surveys proper with analyses of reserve money will be argued briefly.

III. MONEY SUPPLY ANALYSIS AND MONETARY FINANCING IN THE OPEN ECONOMY

The simplest type of analysis looks to the money supply for evidence on, and a measure of, inflationary or deflationary tendencies in the economy. Before going on to examine the merits of such an analysis, a preliminary warning must be issued against using money-supply statistics without taking into account the economy's structural changes. It is a fact that the volume of money — and particularly of certain components of it — in certain economies contains within it a steep trend deriving not only from the progress of production but also from the increasing use of monetary payments in some sectors of the economy. Clearly when the growth of the money supply can be explained by such phenomena there are no grounds for speaking of inflationary tendencies. The relevant concept seems to be that of deviations from trend.

While the analysis of fluctuations in the money supply can be useful in a first approximation when examining inflationary tendencies in a closed economy, it does not provide, *taken by itself*, very significant information in most other cases.

Inflationary credit monetization may be said, following Holtrop,⁶ to result in one of two different types of liquid asset accumulation. The first type expresses an adjustment by disinvestment out of inventories in one sector, to an increase in expenditure in another sector. Accumulated stocks, that is, are converted into liquid assets, but the reaction can be only temporary: there will have been no autonomous increase in the desire to hoard and attempts will consequently be made to build up inventories to their previous level.

6. Holtrop, *op. cit.*

Should this be made impossible by controls, the accumulation of liquidities may result in the well-known case of suppressed inflation.

When, however, real or nominal income increases, the additional liquidities resulting from inflationary financing will eventually become needed as transaction balances, liquidity preference remaining the same here too; this "absorptive reaction" may, of course, take the form of increased cash balances in the hands of the public.

The point is, however, that in an open economy the consolidation of the inflationary impulses via an increase in incomes and the consequent absorption of newly created liquidities is not necessarily the rule. Actually, unless controls are applied to imports, sooner or later the reconciliation of *ex-post* savings and investment would be brought about by disinvestment vis-à-vis the foreign sector, that is an import surplus. The upshot seems to be that inflationary financing will fail to result in an increased money supply, unless controls are applied or there is for other reasons an *autonomous* increase in the desire to hold liquid balances. In fact, even without strict controls, the imperfect flexibility of the economy and the inability of imports to satisfy exactly the excessive internal demand will permit or rather force the absorptive reaction to operate within certain limits. Yet it is clear that the increase in the supply of money in the hands of the public will provide a most imperfect and at times misleading indication of the strength and the origin of inflationary financing.

If inflationary financing is not always reflected in an increase in the money supply — within the usual time limits of our observations — it is equally true that an increase in the public's holdings of liquid assets need not mean that an inflationary process is running its course.

The most obvious example consists in the increase in the money supply that may accompany net purchases of foreign exchange by the banking system. One way in which this liquidity surplus of the economy as a whole towards the foreign sector could come about is, for instance, an autonomous increase in the desire to hoard. Imports may drop or exports be stimulated indirectly, or both, but liquid holdings in the hands of the public, that is hoards, could increase without necessarily having depressive effects. In this case the situation could be identified in a relatively direct way; it is more likely, however, that the deflationary impulse would actually operate on both the foreign balance and the rhythm of activity. As the two opposed influences on the volume of liquidities in the hands of the public would then partially offset each other, only a careful study of credit monetization and the net foreign exchange position, *in addition* to the money supply, could be conclusive.

This change in the liquid assets outstanding in the economy as a whole was associated with a cutback in expenditure. However, an excess of purchases of foreign currency over sales by the banking system, leading to an expansion of the money supply, need not be associated with such a "deflationary" condition. It could be associated with a *relatively less inflationary* situation with respect to other countries: in fact such a general formulation also covers the preceding case.⁷ In principle, the differences between a surplus caused by internal deflation and one attributable to inflationary tendencies abroad are merely differences of degree: for the purposes of practical monetary analysis, however, one could also speak of qualitative differences. Clearly, the accumulation of liquid balances in the case of a surplus due to inflation abroad tends to be temporary only if the home country is in equilibrium or is inflating; it *will be* in its interest to accelerate the rhythm of imports so as to moderate the additional inflationary influence of the favorable trade balance.

The case of the country experiencing an increase in the desire to hoard is different in that it *will not be* in the interest of the country to encourage imports, at least until it has overcome the difficulties which the fall in internal spending may cause.

We can therefore expect, according to the origin of the surplus, substantially different lags before seeing a reaction to the emergence of an over-all liquidity surplus (in the form of net purchases of foreign exchange by the banking system).

In yet a third case money supply analysis may prove misleading for judging whether excess expenditure by any one sector of the economy is being financed by inflationary methods: the case is the transmission of inflationary impulses from abroad in the presence, however, of a *deficit* rather than a surplus in the balance of current payments. Inflation abroad need not be reflected in an export surplus towards the inflating country: the inflation in one or more countries may actually cause a deficit in the balance of trade of all other countries (except the direct suppliers of the inflating country), by raising against them the prices of raw materials they buy from common sources of supply (or from the inflating country itself). This was the plight of a number of countries during the Korean scare-buying. In this situation the money supply is likely to rise concurrently with the emergence of the deficit in the countries that are so affected; this simply means that the destruction of liquidities brought

7. It might actually be convenient for certain purposes (mainly balance-of-payments analysis) to redefine inflation and deflation as relative concepts with international trends as a yardstick.

about by the deficit is being offset, and generally more than offset in order to prevent a liquidity shortage. It appears that credit will also have to be expanded following the increase in prices transmitted from the outside, although the drawing into circulation of idle balances may moderate the stringency somewhat. No automatic stabilization effect can be expected until the impulse transmitted from abroad ceases to operate, as monetary severity would only make things worse; there exists, however, the possibility that the government may undertake to subsidize the sale of import goods at unchanged prices and finance the program by increased taxation, as was done in Great Britain and in other countries during World War II.⁸

IV. INFLATIONARY AND DEFLATIONARY IMPULSES AND REACTIONS

Isolating the external accounts is only a first step in the elaboration of a monetary survey; we need to disaggregate further the accounts of the economy so as to be able to interpret them in terms of expenditure decisions of reasonably homogeneous sectors. Just as there can be compensation between an internal inflationary impulse and a deflationary reaction registered in the balance of payments, compensation among domestic sectors may prevent income-expenditure disequilibria from affecting the liquidity of the country as a whole. It would be interesting, nevertheless, to know when such a situation is occurring, because an equilibrium of this type is significantly different from an equilibrium characterized by lack of disturbances. Actually this is an argument against using simple liquidity balances which are mere algebraic sums of the recourse to liquidity-creating sources and the absorption of newly created liquidities. There is hardly any sector for which a liquidity surplus or deficit has an unambiguous meaning: the examples that follow tend to show that not only the household sector, but also the business and the government sectors may present inflationary or deflationary characteristics as a consequence of spending decisions of *other* sectors.

In the last of the cases discussed above, that of imported cost inflation, the expansion of the money supply was due not only to the fact that higher prices necessitated a larger volume of means of payment to finance normal turnover, but also to the necessity of obviating the decrease in business receipts. An import drain is by no means, however, the only possible cause occasioning such a decrease in the receipts of a sector and hence creating the need for

8. See on this point A. J. Brown, *The Great Inflation* (London, 1955), pp. 254, 288.

the sector to finance a liquidity deficit by means of new credit monetization — even though its volume of expenditure may remain unchanged or perhaps contract.⁹

Supposing that an autonomous increase in the desire to hold liquid balances occurs in the household sector (that is, consumers reduce their expenditures below normal), the flow of funds into the business sector will be reduced. It is common to think of the consequences in terms of reduced economic activity, liquidations of stocks and achievement of equilibrium at a lower level. We should expect, therefore, to see the outstanding volume of liquid assets decline until equilibrium is reached, if we abstract from balance-of-payments reactions. In the short run, however, even if we limit ourselves to the problem of the liquidation of inventories and the decrease in new orders to manufacturers, we shall have to recognize that there will be a time lag in which the carrying over of unsold inventories will result in a relative expansion of credit and of deposits. In addition we could also consider that at any moment of time firms will have begun a certain number of investment projects which it will be worthwhile to complete, and which will therefore require the use of new credit to substitute for the unexpectedly reduced possibilities of self-financing.¹

In an expanding economy and provided confidence is not sapped seriously, the over-all result of an increase in the desire to hoard may therefore be an expansion of the money supply, even though the condition of the economy would normally be considered deflationary.

On rather similar lines, we can consider now a second problem concerning the relations of the government's liquidity with that of the other sectors:

To the extent that Government revenue is affected by deflationary conditions . . . the increase in deposits may reflect an improvement in the liquidity position of the private sector matched by a deterioration in that of the Government, in which case the increase in deposits will be balanced by larger bank lending to the Government.²

The act of "inflationary financing" by the government must be classified as a reaction to a deflationary impulse, when we consider the

9. The two examples that follow owe much to the stimulating paper by Dr. Baffi, "Monetary Analysis in Italy," *IMF Staff Papers*, Feb. 1957.

1. A striking example of credit extension due to the carrying over of stocks and to the reduction in the margins available for self-financing was offered by Sweden in 1954: advances expanded by 1200 Million kroner, yet prices were not affected in the least. See: Lundberg and Senneby, "The Dilemma of the New Monetary Policy in Sweden," *Skandinaviska Banken Quarterly Review*, July 1956, p. 83.

2. Baffi, *op. cit.*

rationale of the creation of new liquidities. This definition is probably more helpful from the practical angle than the traditional definition of any government deficit as an autonomous inflationary factor, as it provides a hint for interpreting subsequent developments; these are likely to involve the disappearance of the government's liquidity deficit as soon as the deflationary impulse ceases to operate.

If such then is the variety of developments that can affect the monetary accounts of the country, and if the meaning of any single synthetic measure — liquidity or finance balance — of the income-expenditure disequilibrium of a sector from the viewpoint of economic (not simply of statistical) analysis is so ambiguous, we must look for a solution by studying the interrelations of the various items among themselves and over time. Professor Copeland, and more systematically, Professors Tinbergen and Schouten³ have proposed criteria for recognizing the role — inflationary, deflationary, reflationary and disinflationary — played by any one sector in the economy and its "responsibility" in assuming such a role. The classification is couched in dynamic terms: an inflationary sector is one whose deficit vis-à-vis the rest of the economy is increasing, a deflationary sector one whose surplus is increasing, a reflationary sector one whose surplus is decreasing, a disinflationary sector one whose deficit is decreasing. The problem of "responsibility" — or of initiative — is approached by observing the parallel developments in a sector's balance and in its expenditure and receipts. Thus, active sectors, those in which the spending decisions have originated, will generally be characterized by a negative (positive) correlation between their current expenditures and the corresponding surpluses (deficits). Passive sectors, in which what we have called "reactions" take place, will be recognizable, according to Tinbergen and Schouten, by a positive (negative) correlation between their current receipts and the corresponding surpluses (deficits).

A word of caution is in order at this point: the classification just presented rests on a line of reasoning which, while theoretically consistent, has practical implications which make it a somewhat delicate instrument of analysis. The underlying argument is clearly a type of period analysis in which by definition there is always *ex-post* equilibrium among the various sectors, so that it is not the absolute position of a sector (e.g., liquidity surplus or deficit) that counts, but the

3. M. R. Copeland, *A Study of Moneyflows in the United States*, National Bureau of Economic Research (New York, 1952), chap. 12, sec. 5; Tinbergen and Schouten, "National Income Accounts as a Means of Currency Analysis," *International Economic Papers*, No. 5, 1955.

change in such a position from one period to the next. Thus a sector running a persistent liquidity surplus of constant amount will not be considered deflationary, as the rest of the economy is supposed to have adjusted to the original deflationary impulse in the period in which it made its appearance and to have continued in its compensatory position ever since. Sometimes, however, such a situation may not justify a lack of concern with monetary developments in the economy and may instead require positive corrective steps, even though equilibrium formally obtains. Whether this will be the case will depend on the nature of the adjustment that has taken place; it is easy to think of cases in which such an adjustment may have undesirable effects on the structure of the economy, may be practicable only for a limited period, or may finally build up dangerous potential disequilibria. Fitting examples would be respectively the adjustments in the structure of production required by a prolonged substitution of compensatory government spending for deficient private demand, the possibility of exhausting foreign exchange resources by financing a persistent balance-of-payments deficit, and finally the creation of potentially destabilizing excess liquidity in the hands of the public by monetary policy measures.

V. THE PROBLEM OF SECTORING

We have been examining a type of analysis of intersector financing that requires a subdivision of the economy according to the criterion of homogeneity of behavior. Current surveys normally present a government sector, a private sector and a foreign sector, the latter sometimes appearing only implicitly, as a residual: others push the sectoring further, segregating central from local governments, isolating insurance and pension funds, introducing the capital market as a separate factor.

The next desirable step, an essential one in fact, is the subdivision of the private sector into a business sector and a household sector, since the behavior of these sectors cannot be considered homogeneous enough in advanced countries. But here statistical problems appear to have been the stumbling block in many countries.⁴

When we come to bank assets other than current account advances we would normally find that financial obligations embodied in securities are recorded as loans to the original issuer: this fails to

4. Recent official documents confirm the impression, however, that the difficulties presented by the identification of bank deposits and bank loans as pertaining to households or business firms are not insuperable. See in particular the 1956 report of the Bank for International Settlements (Basle, 1957) where the analysis is carried out for England, France and Germany.

reveal the source from which they have actually been acquired, and which is therefore responsible for the creation of new liquidities. This could result in misleading conclusions if the original issuer belonged to one sector and the present borrower to another. This would be the case, for instance, if government securities or foreign securities were turned in by the public to money-creating institutions. An ingenious solution to this difficulty has been adopted by the *Nederlandsche Bank* and the *Bank of Belgium*, which have introduced a sector "capital market and sundry items," to which these operations are attributed.

Where the capital market appears, as in our model, simply as a component of the finance balance, a particularly strong reason is added for wanting to separate the household sector from the business sector: if this is not done, a significant volume of transactions in stocks and industrial bonds goes unrecorded.

VI. A SIMPLE MODEL OF MONETARY SURVEY

The above discussion should have clarified many of the issues involved in integrating monetary analysis with the income-expenditure approach. In the simple model of Table I we attempt to provide a framework in which the interrelations we have been discussing can be made explicit. The model is necessarily quite schematic and thus skirts many of the difficult problems of definition that arise in the practical construction of a survey; we shall, however, comment briefly on the significance of the items appearing in the table.

It will be noticed that the foreign sector is not included explicitly in the model: its liquidity and finance balances are identical, of course, and of opposite sign to those of the domestic economy. This is because the liquidity balance for the country as a whole will be equal definitionally to the sum of the net "equilibrating" items in the balance of payments, that is the net change in foreign exchange reserves and in other short-term claims abroad. Similarly, the finance balance for the country as a whole expresses the net change in total indebtedness (including long-term claims) towards the rest of the world, that is, the balance of payments on current account.

Attention should not be concentrated, however, on the total balances, but rather on the operations performed in the various sectors: for instance, capital movements on the domestic market will clearly cancel out for the country as a whole, but are quite relevant for the analysis of intersector transactions.

External autonomous capital movements are an admittedly vague item from the statistical viewpoint, but a perfectly definite one from

TABLE I

SOURCE AND USES OF FUNDS	SECTORS			
	TOTAL Domestic Economy	Government	Business Enterprises	Households
(1) INCOME	360	40	130	190
(2) EXPENDITURE	366	49	167	150
(3) FINANCE BALANCE (= 1 - 2 = 4 + 5 + 6) expressing an income ex- penditure disequilibrium financed by:	-6	-9	-37	40
(4) RECOURSE TO DOMESTIC CAPITAL MARKET AND FINANCIAL INSTITU- TIONS OTHER THAN BANKS (net borrowing)	X ¹	-5	-10	15
(5) EXTERNAL AUTONOMOUS CAPITAL MOVEMENTS (net imports)	-8	-3	-10	5
(6) LIQUIDITY BALANCE (= 3 - 4 - 5 = 7 + 8) expressing a change in liquidity resulting from:	2	-1	-17	20
(7) RECOURSE TO THE BANKING SYS- TEM AND ISSUE OF SECONDARY LIQUID ASSETS (increase in in- debtedness)	-6	-1	-15	10
(8) ABSORPTION (+) or DRAWING DOWN (-) OF LIQUID ASSETS	8	X ²	-2	10

1. Equals zero by definition.

2. Equals zero by definition, inasmuch as any accumulation of liquid funds by the gov-
ernment is automatically used to decrease its short-term indebtedness.

the conceptual viewpoint, stemming from the well-known discussions of exchange equilibrium by Professor Nurkse, Mrs. Robinson and others. Such capital movements must also satisfy the criterion of not giving rise to liquid claims on the importing country. In practice such capital movements may often be considered identical to long-term capital exports and imports.

Theoretical and institutional considerations, not accounting conventions, should determine which changes in indebtedness also modify the liquidity of a sector. From this viewpoint, the relevant criterion should be that liquidity items have a more direct influence on spending decisions and are more immediately affected by other sectors' spending decisions. Also, a sector has considerably greater freedom with respect to the items which enter the liquidity balance than with respect to the remaining items of the finance balance. These cannot be easily manipulated without the co-operation of another sector.

The items in the bottom row of Table I, indicating changes in

holdings of liquid assets, now deserve a closer look. Such changes in gross liquidity (as distinct from changes *net* of variations in indebtedness) have an autonomous significance, but this is unambiguous only on condition that changes in prices and output be, at the most, moderate. If, instead, the developments that lead to variations in liquidity are associated with substantial movements in the price level and in the volume of output, the interpretative and prognostic value of variations in liquid asset holdings is considerably reduced. To obtain significant indications, changes in gross liquidity must then be corrected to allow for the effects of the increased monetary requirements implied by a higher money income (because of increases in prices and/or output).

One must be fully aware of the dangers involved in estimating such monetary requirements on the basis of an average ratio chosen more or less arbitrarily (usually relying on a period thought of as "normal"). It would be even worse, however, to refrain, for fear of imprecision from trying to estimate changes in liquidity in this second, more meaningful sense. In particular, a tendency towards an abnormal liquidity build-up, or, conversely, a strain on the liquidity of the economy, can be recognized only on condition that such estimates are carried out. It is the residual obtained by subtracting from changes in liquid asset holdings the increased absorption due to the factors just mentioned that measures the practically relevant variations in liquidity. Moreover, only such an analysis permits a full appreciation of the role played by hoarding and dishoarding.

From a more sophisticated study of liquidity, monetary and fiscal policy can thus acquire further insights of importance for the choice of the appropriate corrective instruments in inflationary and deflationary situations.

VII. RESERVE MONEY ANALYSIS: THE MODEL

The framework is, however, not yet complete: the decentralized organization of the money-creating system in most countries makes it desirable that a second set of accounts be used to supplement the monetary survey we have been examining. If we consider that at any given moment the banking system can exercise a certain measure of discretion in the creation of money, within the limits dictated by its reserve position, we can easily appreciate the significance of analyzing the developments that affect precisely this element, the liquidity of the banking system. The residual item that such an analysis would yield, the free reserves, should give a rough measure of the leeway which money-creating institutions have for the expan-

sion of credit. There is hardly any need to point out that this presupposes the adherence by banks to certain liquidity standards, either self-imposed or legal ones; the standards ought to result clearly from the analysis, in the form of an item that isolated a part of the system's liquid assets as unavailable for any other purpose than that of liquidity reserve.

It is not easy, however, to be precise, as to which liquid assets should be taken into account in the analysis of reserve money. Most frequently the legal reserve requirements refer to cash holdings — cash in hand and deposits at the central bank. No obvious guidance exists in case there are no legal obligations of this sort for the banks, but even where there are such obligations the relevant liquidity variables from the viewpoint of bank behavior do not necessarily coincide with the legal reserves. It would have been clearly misleading, for instance, to exclude government securities from an analysis of reserve money for the United States in the years up to 1951, inasmuch as they enjoyed a ready and stable market at the Federal Reserve Banks. The money-creating potential of the banking system must clearly take into account also the secondary reserves which the banks would not hesitate (and would be able) to liquidate in case the demand for credit strained their available free reserves. Indeed in recent years monetary policy in some countries seems to have shifted its emphasis from regulation of cash reserves to regulation of the over-all liquidity of the banking system: the United States has certainly learned the lesson from bitter experience, and Great Britain showed the importance it attributed to this factor by using as a major instrument of monetary control drastic reduction of the banks' liquid assets (as in the 1951 funding of Treasury bills and in successive steps).⁵

In general terms, factors affecting bank reserves could be presented as follows:

TABLE II

- (1) CHANGES IN EXTERNAL RESERVES (net purchases = +)
- (2) CHANGES IN CURRENCY IN CIRCULATION (decrease = +)
- (3) OPEN MARKET OPERATIONS BY THE CENTRAL BANK (net purchases = +)
- (4) REDISCOUNT AT CENTRAL BANK AND ADVANCES BY CENTRAL BANK TO COMMERCIAL BANKS AND PRIVATE CUSTOMERS (increase = +)
- (5) MONETARY OPERATIONS OF THE TREASURY, consisting in the net transfer (absorption) of liquid resources by the government to (from) other sectors, including commercial banks (net transfer = +)

5. See, however, E. Nevin, Comments in "Monetary policy: A symposium," *Oxford University Institute of Statistics Bulletin*, Aug. 1952, p. 297 for a skeptical view of the liquidity doctrine. It is noteworthy that in 1955 the Governor of the Bank of England explicitly asked the banks to observe the 30 per cent liquidity

- (6) CHANGES IN LIQUID RESOURCES REQUIRED AS LEGAL RESERVES OR CUSTOM-ARILY HELD, due to
- (a) Increased liabilities (increase = -)
 - (b) Changes in percentage requirements (increase = -)
-
- (7) CHANGES IN THE BANKS' FREE RESERVES, obtained as the algebraic sum of the items above.
-

VIII. RESERVE MONEY ANALYSIS: INTERPRETATION OF THE ACCOUNTS

Some of the data entering reserve money analysis are sometimes obtainable from the monetary surveys themselves, but the focus of attention is different. In the first place, the analysis of bank liquidity provides a guide to the supply side of the money creation mechanism which traces the ultimate sources of liquidity beyond the point normally apparent in monetary surveys. In the second place, the influence of policy measures such as changes in reserve requirements, the establishment of rediscount ceilings, etc., is accorded a prominent position. Finally — and this is the central argument for reserve money analysis — studying bank liquidity means observing in a sense the preparatory stage of the situation recorded in a monetary survey; it is particularly useful, therefore, in evaluating in advance the probable effectiveness of restrictive monetary measures. It may also help identify any offsetting factors that may frustrate the action of the monetary authorities. It is not enough to identify the sources of the disequilibria that result in inflationary tendencies in the economy: it is also necessary to have a clear vision of the enabling factors. Moreover, certain operations of a monetary character performed by the government, such as debt conversion operations, may involve no net deficit for the government itself, and go unrecorded in monetary surveys. Yet the consequent change in the liquidity of the banking system is surely significant.

When we consider the continuous adjustment of the monetary system to autonomous disturbances, and the difficulty which is thus introduced in a type of analysis based upon static observations, we can appreciate another aspect of the contribution which reserve money analysis can bring to liquidity analysis in general. If there is a functional relation of the sort we discussed between bank liquidity and expansion of credit, we can say that joining the two forms of analysis adds a time dimension to our observations. We must be clear about the conditions under which this is most nearly the case: central bank action and government action must be close to neutral: see H. G. Johnson, "The Revival of Monetary Policy in Britain," *Three Banks' Review*, June 1956, p. 6.

trality and we must assume as an economic background a period of substantially even growth.

Given these perhaps not unduly restrictive assumptions, we face the general problem of determining the time path of the various types of reactions to autonomous changes. This involves a structural analysis of some characteristics of the economy and of the monetary and banking system in particular, which have been ably discussed, with an attempt at empirical estimation, by Polak and White. The problem examined by the two authors relates specifically to the time path of a balance-of-payments drain in the presence of income expansion, but is clearly susceptible of wider applications.⁶ The structural factors they discuss are partly of a very familiar nature and determine the shifts of the supply and demand curve for "money" (one could better say "liquid assets"): they are, of course, the income velocity of money, the marginal propensity to import and the reserve ratio of the banking system (or, evidently, the type of liquidity ratio we have indicated, if this is more relevant). The rate of change of the total money supply will depend upon the elasticity of demand for money (as a function of the rate of interest) on the part of the public, and upon the elasticity of supply of money (as a function of the same variable) on the part of the banks. While the Polak-White model is presented in a rigorous form, designed to be tested by econometric methods, looser formulations, of course, are also useful, and are perhaps more widely applicable at the present stage of monetary analysis. This is substantially what we have been trying to achieve and most of the two authors' key variables will be seen to have appeared in the course of our discussion. Leaving aside the income velocity of money and the reserve ratio, of which enough has been said, we can remember, for instance, the discussion of inflationary or deflationary conditions as a factor influencing the propensity to import of the country under consideration.

The elasticity of demand for money comes in at the stage when the general trend of the economy is considered. For instance — to remain within Polak and White's example — boom conditions have definite implications as to the sensitivity of the demand for loanable funds to movements in the rate of interest caused by a balance-of-payments drain. Finally, the attitude of the monetary authorities and the institutional factors affecting the pattern of interest rates and the availability of credit cannot be forgotten, when the elasticity of supply of money is considered.

6. J. J. Polak, and W. H. White, "The Effect of Income Expansion on the Quantity of Money," *IMF Staff Papers*, Aug. 1955.

THE FUNCTION OF WAGE POLICY: THE AUSTRALIAN EXPERIENCE*

By J. E. ISAAC

Introduction, 115. — I. Institutional features and principles of wage-fixing in Australia, 116. — II. Wages and prices since 1945, 120. — III. Comparison of award and actual rates, 125. — IV. Function of wage policy under full employment, 127. — V. Conclusion, 136.

The essential features of a national wage policy are, one, a high degree of centralization in wage determination. Two, a large measure of control by the central authority over changes in the general level of wages and the wage structure. And three, movements in wages depending largely on national social and economic considerations rather than narrowly on the state of particular firms and industries.

Subject to these general requirements, there may be substantial variations in the manner of administering national wage policy. For example, the government may enter directly into the control of wages, as in the Netherlands¹; it may play an informal but important part by close consultation with trade union leaders, as in Norway²; its informal influence may be less important, as in Sweden³; or, as in the case of the federal government in Australia, it may play no part in the control of wages other than to set up independent tribunals to settle industrial disputes. There may be important differences in procedure and in the principles of wage determination. And so on. All these variations reflect differences in the structure of trade unions and employers' associations; in the structure and constitutional framework of government; in the structure of industry and urgency of national economic problems; and in ideology and economic objectives.

A method of administering wage policy has been operating in Australia for some twenty-five years. It was first used in the depres-

*The substance of this paper was presented to Professor John T. Dunlop's Labor Seminar at Harvard. I am indebted to Professor Dunlop and to members of his Seminar for helpful comments.

1. See B. Zoetewij, "National Wage Policy: the Experience of the Netherlands," *International Labour Review*, LXXI (Feb. 1955), 148.

2. W. M. Leiserson, "Wages in a Controlled Economy" (Unpublished Ph.D. thesis, Harvard University, 1955).

3. W. Galenson, "Comparative Labor Movements," (New York: Prentice-Hall, 1952), p. 165.

sion of the thirties and it has since been applied in full employment, in war and in peace. This method of implementing wage policy did not result from deliberate legislative design, but has evolved in a federal system under economic and institutional pressures from piecemeal attempts to deal with industrial disputes through compulsory arbitration.

The place of a national wage policy under full employment is a controversial one. Some writers⁴ have stressed the need and importance of such a policy. Some⁵ have emphasized the difficulties of enforcing such a policy. And others⁶ have cast doubt on the need or advisability of a national wage policy. Australian experience in the last ten years does not help much to resolve these differences. In fact, the doubling of the price level between 1945 and 1952 raises the question whether anything deserving of the title "national wage policy" did really exist in Australia. But is price stability the hallmark of a "successful" wage policy? Or does wage policy serve other purposes? These questions are posed by the Australian experience and they cannot fairly be answered outside the context of the particular institutional and economic environment in which wage policy operates. There is need for special caution in framing criteria for wage policy which are presumed to have universal application.

This paper will deal first, with the institutional features and principles of wage-fixing which provide the necessary basis for national wage policy in Australia; next, with the course of wages and prices since 1945; and finally, with some of the problems arising from wage policy in this period, especially with the function of wage policy.

I

The determination of wage rates in Australia is in the hands of a large number of tribunals, constituted under state and federal law. These tribunals vary in name, form, procedure, and jurisdiction. It is not proposed here to describe in detail this complicated wage-

4. For example, W. H. Beveridge, "Full Employment in a Free Society," pp. 198 *et seq.* A. Flanders, "Wage Policy in Full Employment in Britain," *Bulletin of the Oxford University Institute of Statistics*, Vol. 12 (July-Aug. 1950). A. P. Lerner, "Economics of Employment," chaps. 14 and 15.

5. H. W. Singer, "Wage Policy in Full Employment," *Economic Journal*, LVII (Dec. 1947), 438. R. Meidner, and E. Lundberg, in "Wages Policy Under Full Employment," ed. R. Turvey, (London: Hodge, 1952), chaps. 2 and 4.

6. E. H. Phelps Brown and B. C. Roberts, "Wages Policy in Great Britain," *Lloyds Bank Review*, Jan. 1952. D. T. Jack, "Full Employment in Retrospect," *Economic Journal*, LXII (Dec. 1952), 731. W. A. Morton, "Trade Unionism, Full Employment and Inflation," *American Economic Review*, XL (Mar. 1950), 13.

fixing machinery.⁷ It is important to note, however, that over the years there has emerged a distinct leader from among the different tribunals. This is the federal tribunal known as the Commonwealth Court of Conciliation and Arbitration.⁸ The preponderant authority of the Court rests partly on the fact that half of all wage earners, spread over the whole country, work under its awards; and partly from the prestige which the Court has acquired over time.

This tacit centralization of decision-making in wage-fixing has been encouraged by the centralized structure of unions representing nearly 60 per cent of all wage and salary earners, and employers' organizations. These central representatives of labor and employers do not possess the coercive powers over their constituents of their Scandinavian counterparts.⁹ But the fact that wage policy is pronounced by the Court and does not necessarily carry the approval of central leaders largely removes the need for any real disciplinary powers. The main responsibility of the central representatives is to formulate the general claims of their respective constituents as strongly as possible, not to enforce the decision of the Court. Unfavorable decisions do not reflect on the central representatives who are always in a position to throw the full weight of any dissatisfaction onto the Court. This, paradoxical as it may seem, has the effect, at least among unions, of promoting a greater feeling of solidarity; and it usually helps to make the Court's rulings more widely accepted, even if only as a "temporary measure, while the next move is being planned."

There is, however, an important qualification to be made concerning this kind of tacit centralization through the Court's leadership. Under the Australian federal system, the federal government has no powers to legislate directly on wages, hours, or any other condition of work. But it is empowered to provide the machinery necessary to settle interstate industrial disputes; and this machinery, of which the Court is the key component, fixes wages and other terms of employment in the course of settling disputes. The state governments, however, are free to legislate directly on all these matters for

7. See K. F. Walker, "Industrial Relations in Australia" (Harvard University Press, 1956), especially chaps. II and VIII.

8. To be referred to henceforth as the "Court." Since this paper was written the Commonwealth Arbitration Court has been reconstituted into two bodies. The determination of wage rates, hours of work, etc. (the arbitral functions) are now carried out by the Commonwealth Arbitration and Conciliation Commission. The interpretation and enforcement of awards, regulation of unions and employers' organizations, etc. (the judicial and administrative functions) are now in the hands of the Commonwealth Industrial Court.

9. See Galenson, *op. cit.*, pp. 128 *et seq.*

all workers who come under the jurisdiction of the state tribunals. A state government may thus order its tribunal to act, within its jurisdiction, contrary to the rulings of the Court. This has happened recently in connection with the system of automatic cost-of-living adjustments; and, as will be discussed below, the possible intrusion of state governments, as a consequence of political pressure, into wage policy has an important bearing on the function of the Court's wage policy.

The impact of the Court's decisions on the general wage level and the wage structure operates through a peculiar classification of wage rates which it has devised and which has been universally accepted. Every wage rate is in two parts — the basic wage and the margin. Originally, the basic wage was intended as the minimum wage for the "purely" unskilled worker, and in this sense it became a national minimum wage. Until 1953 it was the general practice to preserve the real value of this wage by automatic quarterly adjustments in proportion to changes in the cost-of-living index. A margin is added to the basic wage for special requirements of the job — skill and responsibility, danger and dirt, etc. However, the purely unskilled worker has gradually become extinct and nearly every wage now has a marginal component. The basic wage is, therefore, more than a national minimum wage. More significantly, it is the foundational element of all wage rates. A change in the basic wage is a change *par excellence* in the general level of wages because it affects every wage rate directly.

The impact of change in the basic wage is, of course, greater for those wage rates in which the basic wage forms a larger proportion. And hence a change in the general wage level through a variation in the basic wage produces a change in the relative wage rates of skilled and unskilled. These differentials may be manipulated more directly through a change in the general¹ level of margins. In this connection the key margin is that of the fitter in the engineering industry. A variation in this margin provides the cue for a change in all other margins. Thus any undesirable "distortions" in the relative wage rates of skilled and unskilled may be corrected by a direct alteration in the general level of margins.

The awards of the Court (and those of other tribunals) usually fix minimum wage rates, and, except in periods of national emergency, there are no legal restrictions against employers paying more than the rates awarded by the Court. The ability of the Court to control

1. Particular margins being determined by reference to the particular requirements of jobs.

the general wage level and the wage structure, depends, of course, on the extent to which its awards are observed as minima as well as maxima. It will be understood that under conditions of full employment a strong pressure will exist for actual wage rates to be higher than award rates. The extent of these overaward payments² in the postwar years will be examined below, though for the moment it may be noted that the extent of overaward payments has not been large enough to deny the importance of the Court's role in wage movements.

Regarding the third requirement of a national wage policy, that it should be formulated with respect to national considerations, it should be emphasized that the Court is not empowered to formulate and administer a national wage policy as such. As has been noted earlier, its constitutional basis is much narrower and is confined to the prevention and settlement of interstate industrial disputes. Nevertheless, the manner in which the Court has conducted its hearings, the evidence it has sought on national economic matters, the principles it has adopted for the basic wage, the margin of the fitter, and standard hours of work, suggest that it has all but in name assumed the role of formulating national wage policy.

In the early days the determination of the basic wage centered almost entirely around the social question of the minimum needs of a family unit. But since the depression of the thirties the Court has explicitly introduced national economic considerations as its main criteria for basic wage adjustment. It has regarded a claim for a change in the basic wage, in the fitter's margin, and in standard hours as requiring a comprehensive review of the economy. Consequently, disputes on these matters are not ordinary disputes. They are created in order to enable the Court to hear elaborate evidence on the state of the economy and to announce its wage policy.

This outline of the main features of wage policy in Australia shows that in a comparatively small, simple, and highly unionized economy, with most wage earners concentrated in a few large cities,³ the existence of centralized unions and employers' organizations, the

2. The term "wage drift" popularized by the Scandinavians signifies much more than the payment of wage rates above agreed or awarded levels. It includes the effects on earnings of a change in the distribution of labor in the direction of higher paid industries and occupations, overtime work, an increase in piece rate earnings, etc., as well as the payment of rates above agreed levels. For our purposes, we are mainly interested in the extent to which the Court's wage rates are also the effective wage rates in practice, indicating the extent of the control exercised by the Court over wage rates.

3. There are less than three million wage and salary earners in Australia. More than one-third of the population is concentrated in Sydney and Melbourne, and more than half lives in the six capital cities.

leadership of the Court and the high degree of formal interdependence⁴ in wage rates through the determination of two key wage rates — the basic wage and the fitter's margin — provide the necessary basis for a national wage policy without complete legislative and administrative provisions for its operation. There are, however, a number of weaknesses in the procedure and the composition of the Court⁵; and, as will be discussed later, the separation of the Court from other instruments of economic policy and the possible challenge to its leadership by state legislatures, impose an important limitation on the function of the Court's wage policy.

II

The postwar decisions of the Court may be followed in Table I, which also shows in contrast the wage and price movements in the

TABLE I

June to June	AUSTRALIA		UNITED STATES		UNITED KINGDOM	
	Wages ¹	Prices	Wages ²	Prices	Wages ³	Prices
1945-46	2	2	8	3		
1946-47	8	2	14	18		
1947-48	23	9	8	9	5	10
1948-49	9	10	6	-1	3	1
1949-50	9	9	2	0	1	2
1950-51	23	20	9	9	8	10
1951-52	19	20	4	2	8	11
1952-53	7	4	6	1	5	3
1953-54	1	1	3	1	5	1
1954-55	2	1	2	-1	5	3
1955-56	5	7	5	2		

Sources: Commonwealth Bank of Australia, *Statistical Bulletin*; United Kingdom Central Statistical Office, *Monthly Digest of Statistics*; U.S. Department of Labor, *Monthly Labor Review*.

1. General average hourly male wage rates.

2. Gross average hourly earnings in manufacturing.

3. General average weekly male wage rates.

United Kingdom and the United States. The 1946-47 increase in wages in Australia may partly be ascribed to the 7 per cent rise in the real basic wage in December 1946. The big jump in 1947-48 resulted from the reduction in the standard working week from 44 to 40 hours in January 1948⁶ and also from the increase in the general level of

4. See J. T. Dunlop and M. Rothbaum, "International Comparisons of Wage Structures," *International Labour Review*, LXXI (Apr. 1955), 350.

5. See J. E. Isaac, "The Basic Wage and Standard Hours Inquiry in Australia," *International Labour Review*, LXIX (June 1954), 570.

6. See O. de R. Foenander, "The 40-Hours Case and the Change in Standard Hours in Australian Industry," *International Labour Review*, LVIII (Dec. 1948), 717.

margins in this period. The smaller increases in 1948-50 were largely due to the automatic cost-of-living adjustments which appear to have been fairly substantial in this period. Between February 1948, immediately following the reduction in the standard working week, and November 1950, just before the 14 per cent increase in the real basic wage, cost-of-living adjustments alone raised the basic wage by 28 per cent. Early in this period, the federal government abandoned price control and in September 1949 the exchange rate was depreciated against the United States dollar.

The next two years, 1950-52, showed a sharp rise in wages and prices resulting from the combined effects of a 14 per cent increase in the real basic wage in December 1950 and the subsequent cost-of-living adjustments. In this period, the price of wool doubled in the season following the outbreak of the Korean War and fell back to its former level in the next. This was followed by a period of comparative stability in wages and prices up to 1955. In September 1953 the Court abandoned the automatic cost-of-living adjustment system, but as prices remained stable this was of nominal significance until 1955-56 when a renewal of price increases following a general increase in margins and the beginnings of a domestic inflation produced a strong demand for a return to automatic cost-of-living adjustments. The Court rejected this demand but raised the basic wage by 4 per cent, or about half the amount necessary to offset fully the loss in the real basic wage since the automatic system was discontinued. In the ten years ending in June 1955 the general level of hourly male real wage rates rose by 25 per cent. But this rise involved an increase of 163 per cent in money wage rates.

It is interesting to notice (Table II) that with all these wage and price changes, the share of wages and salaries in national income has emerged at a higher figure than its prewar level. Furthermore, these figures may understate somewhat the rise in the share of wages, partly because the 1938-39 figure was abnormally high on account of a depression farm income resulting from a severe drought; and partly because salaries lagged behind wages in this period. Between 1946-47 and 1950-51 labor's share shows a downward trend. In this period the rise in export prices improved the relative position of the farmer. This is reflected in the rise in the ratio of farmers' income to wages and salaries. Wages and salaries also declined in relation to nonfarm profits in 1950-51. The wool boom reduced labor's share to its lowest level in 1950-51 but the sharp rise in wages and the subsequent fall in wool prices more than restored labor's share.

The sharp fall in the ratio of nonfarm profits to wages and salaries

from the peak in 1950-51 to 1952-53 shows the effect of the large increase in wages which continued right through a short period of recession and tight market conditions in 1951-53. The ratio rises somewhat in 1953-55 following the recovery and comparative stability in wages, but it is well below its prewar level, suggesting that the higher share of labor has been at the expense of nonfarm profits. A more detailed analysis of these figures would show that the higher share of labor in the postwar period has been largely at the expense of the rent and interest component of nonfarm profits.

TABLE II
WAGES AND SALARIES, FARMERS' INCOME, NONFARM PROFITS,¹
AND NATIONAL INCOME

	Wages and Salaries as % of National Income	Farmers' Income as % of Wages and Salaries	Nonfarm Profits as % of Wages and Salaries
1938-39	56	10	62
1939-40	53	19	63
1946-47	56	22	53
1947-48	52	38	53
1948-49	54	30	52
1949-50	52	37	53
1950-51	48	50	57
1951-52	58	23	49
1952-53	57	28	47
1953-54	57	24	50
1954-55	58	19	52
1955-56	59	16	51

Source: Commonwealth Bureau of Census and Statistics, *National Income and Expenditure, 1955-56*.

1. Company income, income of unincorporated businesses, professions, etc., and net rent and interest. A small proportion of farm profits is included in company income.

The main economic factors behind these wage and price increases have been discussed elsewhere.⁷ Here it will be sufficient to point out a few facts, some of which are confirmed in Table III. The Australian economy is vulnerable to fluctuations in export income, which contributes between 15 and 20 per cent of Gross National Product, and to fluctuations in capital imports. Exports are highly specialized in primary products with wool alone accounting for half, and sometimes more, of total export income. Through most of this period, capital formation and immigration were at a high level. Unemployment was negligible, except for the short period of recession,

7. J. E. Isaac, "The Claim for a £10 Basic Wage in Australia," *International Labour Review*, LXIII (Feb. 1951), 149 and *idem*, "The Basic Wage and the Standard Hours Inquiry in Australia," *op. cit.* D. C. Rowan, "The Monetary Problems of a Dependent Economy," Banca Nazionale del Lavoro, *Quarterly Review*, Dec. 1954. R. I. Downing, "The Australian Economy, March 1956," *Economic Record*, XXXII (May 1956), 1.

TABLE III

SOME INDICATORS OF THE STATE OF THE AUSTRALIAN ECONOMY

	Number Receiving Unemployment Benefits (Average weekly figures) ¹	Ratio of Export Prices to Import Prices (1936-37/ 1938-39 = 100) ²	International Reserves at 1945-46 Import Prices (June 30) ³	Public Authorities' Annual Deficits as Percentage of Public Authorities' Outlay ⁴	Gross Capital Formation as Percentage of Gross National Product ⁴	Rate of Net Immigration ¹ (Calendar years beginning 1946)
1945-46	5,514	74	225	—	—	-0.20
1946-47	9,371	70	197	6	—	0.14
1947-48	3,939	105	191	-1	—	0.72
1948-49	1,573	117	318	1	19	1.90
1949-50	13,057	125	414	12	23	1.86
1950-51	784	177	435	13	24	1.32
1951-52	2,391	111	180	8	29	1.09
1952-53	29,510	122	286	12	24	0.50
1953-54	13,812	123	297	6	25	0.76
1954-55	3,871	110	220	9	26	1.06

1. Commonwealth Bureau of Census and Statistics, *Quarterly Summary of Australian Statistics*.2. Commonwealth Bank of Australia, *Statistical Bulletin*.3. Commonwealth Bank of Australia, *Annual Report*, June 1955.4. Commonwealth Bureau of Census and Statistics, *National Income and Expenditure, 1954-55*.

1952-53, following the Korean boom. Since 1952 a chronic balance-of-payments deficit seems to have developed, requiring the use of import restrictions.⁸ It should be mentioned also that in the crucial phases of inflation the government's economic policy was, for political reasons, inadequate to the task with which it was confronted.⁹

The general wage movements discussed above conceal the impact of basic wage changes on the wage structure, more particularly the relationship between skilled and unskilled wage rates.¹ By international standards, percentage differentials for skill have always been on the low side in Australia.² Several factors may have been working towards small differentials but not least among them has been the emphasis of unions, supported in the early stages of wage-fixation by the Court, on the minimum needs of the unskilled worker.

In 1947-48 the slight compression of the war and early postwar period was eliminated by widespread adjustment in margins. However, as was pointed out above, the following years saw a substantial increase in the basic wage component both through a real increase and cost-of-living adjustments. Some idea of the compression may be obtained from the following (Melbourne) ratios of the engineering fitter's margin to the wage rate of the fitter's assistant:

December 31	Per Cent
1938	23
1944	22
1946	20
1947	23
1948	20
1953	11
1954	20

Source: Commonwealth Bureau of Census and Statistics, *Labour Reports*.

By 1952 fresh demands were being made by the unions for an increase in margins in order to restore the 1947 relative differentials, but the Court³ refused to make any adjustment on the grounds of the

8. See E. Lundberg and M. Hill, "Australia's Long-Term Balance of Payments Problems," *Economic Record*, XXXII (May 1956), 28.

9. See Colin Clark, "The Budget and the Basic Wage," *Economic Record*, XXVI (Dec. 1950), 179.

1. Because of the centralized system of wage-fixing and the interdependence of wage rates through the basic wage and the fitter's margin, interregional and interindustrial differentials have generally been small and largely reflect relative job requirements.

2. See A. G. B. Fisher, "Education and Relative Wage Rates," *International Labour Review*, XXV (June 1932), 742; *idem*, "Changing Wage Structures," *International Labour Review*, LXXIII (March 1956).

3. More strictly, the conciliation commissioner charged with the settlement of the dispute on margins in the metal trades. An appeal from his decision later brought the matter up for the Court's consideration.

inflationary situation. And it was not until December 1954 that margins were raised to correct the narrowed differentials.

III

The wage rates referred to above are mostly award rates, and hence minima. The question arises as to what extent actual wage rates departed from these award rates. This would provide some indication of the degree of influence exercised by the Court on the prevailing level of wage rates. The available statistics do not permit a precise answer, but some idea may be obtained from a comparison of changes in average weekly wage rates and average weekly earnings (Table IV). The latter includes salaries (comprising between

TABLE IV

WEEKLY WAGE RATES, WEEKLY EARNINGS, ¹ AND THE WAGE-DRIFT				
	Weekly Male Wage Rates (All Industries)	Average Weekly Earnings (All Industries)	Wage Drift	Wage-Drift as % of Average Weekly Earnings
	Annual Percentage Increases	Annual Percentage Increases	Annual Percentage Increases	Increases
1946-47	5.9	3.7	-2.9	-80
1947-48	10.0	12.2	2.2	18
1948-49	12.3	13.6	1.3	10
1949-50	8.9	9.7	0.8	9
1950-51	20.1	20.1	0	0
1951-52	22.2	23.1	0.9	4
1952-53	10.6	9.6	-1.0	-10
1953-54	3.3	4.8	1.5	31
1954-55	1.3	4.4	3.1	70
1955-56	4.3	6.6	2.3	35

Source: Commonwealth Bureau of Census and Statistics, *Labour Report*, and *Monthly Review of Business Statistics*.

1. Wage and salary earnings, seasonally adjusted, average earnings per male unit employed. (Male units represent total male employment plus a proportion of female employment based on approximate ratio of female to male earnings.)

10 and 15 per cent of the wage and salary bill), which would tend to understate the rise in wage earnings because salaries tended to lag behind wages in this period. However, the discrepancy between rates and earnings (commonly referred to as the "wage-drift") in this period has not been entirely due to overaward payments. Variations in overtime work (paid usually at the rate of time-and-half) would produce a discrepancy between changes in rates and earnings. The big wage-drift in 1947-48 and 1948-49 may be explained partly by the large increase in overtime work following the introduction of the 40-hour week.⁴ The effect of the 1952 recession on overtime work is

4. "Guaranteed" overtime work as a disguised form of overaward payment was not unknown but its incidence was probably small.

reflected in the smaller increase in earnings as compared with award rates. The same factor may have been at work in 1946-47 following the transition from war to peacetime work.

Overaward payments were probably most marked in 1954-55, a year in which award rates rose by only 1 per cent. In the interest of price stability the Court had abandoned cost-of-living adjustments in 1953 and had refused the unions' claim for a general increase in margins. But the existence of excess demand in the face of labor dissatisfaction with the wage structure and the rise in the cost of living defeated the Court's objective for price stability.

Nevertheless, for most of this period the relative extent of overaward payments was not so great as it appears to have been in the United Kingdom, Sweden, and Norway.⁵ This is a little surprising at first sight, considering that the Australian rates were mostly award rates imposed on the parties by arbitration tribunals, and not rates agreed to under collective bargaining. Part of the explanation may be found in the fact that until the end of 1949 wage rates fixed by tribunals were maxima as well as minima.⁶ There were penalties for departing from award rates but these penalties could hardly in themselves account for the smallness of overaward payments, although, of course, they may have set the tone of restraint in the market. A more important explanation, and one which holds after 1949, is that governmental authorities (employing one-quarter of all wage and salary earners) and most large firms (employing perhaps another one-quarter of the labor force) are not in the habit of departing from award rates. This practice would in itself moderate the pressure for overaward payments in the other half of the economy, where perhaps building and construction was the most flagrant source of overaward payments.

However, these explanations do not adequately account for the smallness of the gap between actual and award rates. For if the Court had practiced right through this period the degree of wage restraint required to maintain price stability, it would be very surpris-

5. G. Penrice, "Earnings and Wage Rates, 1945-55," London and Cambridge Bulletin, No. 16, in *The Times Review of Industry*, Dec. 1955. H. A. Turner, "Wages: Industry Rates, Workplace Rates, and the Wage-Drift," *Manchester School*, XXIV (May 1956), 95. Nils Killgren, "Wage Policy in Sweden," *Financial Times Survey of Sweden*, July 1956. M. W. Leiserson, *op. cit.*, p. 150. These international comparisons must not be pressed too far but they do support the impression that the wage-drift for whatever reason has been milder in Australia than in these countries.

6. The Court relied on special Defence Powers to enforce maximum rates. However, by 1949, with the war some years behind, reliance on these powers was of dubious legal validity, and the Court removed the ceiling on its awards.

ing indeed if the departure from award rates had not been much greater. It would appear, therefore, that the Court's minimum wage rates were in most cases high enough to be the maximum wage rates which employers were prepared to pay, the high level of employment notwithstanding. It has been noted that with the more conscious phase of wage restraint after 1953, the departure from award rates appears to have gathered strength.

What the course of wages would have been in the absence of wage determination by compulsory arbitration or a conscious wage policy is impossible to say with complete confidence. However, there are grounds for believing that the Court may have accentuated the rise in wages beyond the levels which otherwise would have prevailed. In the first place, the decisions of the Court were applied on a nationwide basis and not merely in areas where trade union power, or the market situation, or a low-wage-total-cost ratio would probably have secured these increases regardless of wage regulation.

Second, individual employers who might at times have strongly resisted the particular scale of wage increases have tended to accept them passively, especially when effective monetary restraints were absent, because the national scope of the increases leaves the relative positions of employers unaltered.

Third, the existence of sectors, such as building and construction, in which workers have been able to obtain overaward rates, must have built up pressure in the rest of the economy for wage increases; and these in turn have stimulated demands for increases in the overaward sectors in order to maintain differentials. This mutual interaction between award and overaward rates may have tended to intensify wage increases especially as the Court has been influenced by evidence of the extent of overaward payments.⁷

IV

The unrestrained character of Australian wage policy raises the question of the function of wage policy under full employment. Wages, being a significant element of demand and costs, have an important bearing on the level of employment, the balance of payments, the level of prices and the allocation of resources. But the determination of wages is also associated with a particular institutional setting and may be strongly influenced by notions of equity and social justice in the distribution of income. Ideally, it should be the object of wage policy to strike some sort of a balance between the

7. Cf. H. W. Singer, *op. cit.*, p. 447, and H. A. Turner, *op. cit.*, pp. 117 *et seq.*

competing claims of these elements. But in practice circumstances may arise which make it necessary for those formulating wage policy to give primary emphasis to some aspects of wage policy to the neglect of others.

Australian wage policy has had to grapple with the thorny problem of whether it is to be used mainly to meet the broad national objectives of price stability, balance-of-payments equilibrium, economic development, etc., the income distributional consequences being more or less incidental and subordinate to these objectives. Or, whether, in the interest of industrial peace the main focus of wage policy should be more directly on the income distribution question, wage policy acting, if necessary, inconsistently with other objectives.⁸

The Court has not been altogether too sure of itself in this regard. It seems to be torn between narrowly fulfilling its statutory function of maintaining industrial peace on the one hand, and on the other, acting as an arm of economic policy — dismembered though it be from the main body of economic policy. But on the whole, the Court has inclined more to the former, with the inflationary consequences noted above.

Such a stand is not only reasonable but under the circumstances unavoidable for two reasons. First, to depart from the equity standards of the market in the interest of general economic policy would seriously undermine the authority of the Court as a wage-fixing institution. For even if a wage policy is not entirely acceptable to the parties, it will be suffered more readily if it bears certain recognizable imprints of equity. Second, the small contribution which wage policy might make to general economic objectives can be realized only if it is an integral part of general economic policy. The significance of these two points will be apparent in what follows.

First, consider the problem of relative wages. The need for a flexible wage structure, especially in times of full employment, in order to promote the movement of labor to its most urgent uses has been emphasized by a number of writers.⁹ Such a policy is derived from the assumption of a freely working economic system where wage

8. The distributional question looms so large in Australia because of the existence of an important external sector whose prices and profits are largely dependent on the world market. Consequently, changes in the terms of trade and the general wage level produce significant shifts in the distribution of income between wages and profits in this sector.

9. A. P. Lerner, *Economics of Employment*, p. 213. E. Lundberg in R. Turvey (ed.), *Wages Policy Under Full Employment*, (London: Hodge, 1952), p. 8. E. H. Phelps Brown and B. C. Roberts, "Wage Policy in Great Britain," *Lloyds Bank Review*, Jan. 1952, p. 29.

rates are determined in a decentralized manner by the interaction of supply and demand. But it is even doubtful whether under a decentralized system small variations in relative wage rates are in practice effective in reallocating the labor force, particularly where geographical mobility is involved; and, further, there is impressive evidence of the greater importance of job opportunities in distributing the labor force.¹

However, the more centralized the system of wage determination, the more impracticable it is to use discriminatory treatment between industries simply to promote movement, even if it were possible to decide upon an objective basis for discrimination. Consistency becomes a pressing requirement in the determination of wage changes. The formula for standardization which meets the equity notions of the market is that relative wage rates should be based on relative job contents, regardless of the industrial attachment of the jobs.

Australian tribunals have attempted to apply the equity principle in fixing margins, using the margin of the engineering fitter as a yardstick. A change in the fitter's margin usually provides a justification for a general change in margins regardless of industrial incidence. It is true that there are difficulties in applying this principle over the whole range of industries, especially when these vary considerably in nature, market conditions, and location. But it should be remembered that tribunals do not, as a rule, have to construct *ab initio* the whole wage structure based on the equity principle but only to vary wage rates on this principle. It will be conceded that the larger the geographical spread of the work force and the greater the diversity in industrial conditions, the more difficult it becomes to apply this equity principle consistently, and indeed, the more difficult it is to formulate a national wage policy at all.

But in small countries such as the Netherlands, the Scandinavian countries, Australia and New Zealand, the concentration of the working population, especially the industrial work force, not only makes the equity principle workable, but is also a necessary expedient for industrial peace. And as long as there are no serious structural problems, such as a depressed industry with an immobile concentration of labor, there is a good economic case, under conditions of full employment, for standardizing wage rates on the basis of job requirements, throwing the onus for labor mobility on profit differentials rather than on interindustry differentials. The closing down of marginal firms is not only likely to promote mobility more speedily,

1. L. G. Reynolds, *The Structure of Labor Markets*, p. 245.

but will also involve an economically more efficient use of resources.²

In a growing economy such as Australia, where economic development depends on the construction of roads, housing, water conservation facilities, hydro-electric power, etc., it might appear at first sight that considerable emphasis should be given to the wage structure as a means of promoting the movement of labor to these industries. However, to encourage mobility of the existing work force to these socially less attractive industries (because of their distance from the urban centers) would require a considerable distortion in the conventional wage relationships between industries. This is a situation which would not be tolerated in a centralized union organization and would sooner or later be corrected by corresponding increases in the rest of the economy. Some "distortion" in relative wage rates did in fact occur in the form of overaward payments but the need for any serious distortion was obviated by the fact that economic development has been accompanied by immigration, especially from countries with a lower standard of living or with high unemployment,³ thus providing a less discriminating source of labor to the less attractive industries and locations.

However, a difficulty arises in connection with the general level of margins in relation to the basic wage — or broadly speaking, the general premium for skill. For although relative job requirements determine the value of margins, the general premium for skill may be altered as a result of changes in the basic wage component. In Australia, as in other countries,⁴ a considerable compression in the percentage premium for skill takes place in periods of inflation because of the practice of across-the-board increases, these increases being proportionate to the basic wage. Furthermore, the real value of the basic wage has also been increased over time. In fact, between 1946 and 1951, the real value of the basic wage was raised by 20 per cent, which together with cost-of-living adjustments narrowed percentage skill differentials considerably.

The traditional emphasis on the basic wage comes from a concern for those in the lowest wage bracket, but the magnitude of cost-of-living adjustments in this period went a little too far in distorting the conventional differences between skilled and unskilled wage rates, which, as we have noted, have generally been small compared with

2. See B. Zoetewij, *op. cit.*, for an excellent defense of the equity principle. Also, G. Rehn, in R. Turvey, *op. cit.*, chap. 3.

3. See D. Cochrane, "Australian Postwar Immigration," Banca Nazionale del Lavoro, *Quarterly Review*, Mar. 1955.

4. See A. G. B. Fisher, *op. cit.*

those in countries in a similar or more advanced state of economic development. Consequently, by 1952 the unions were demanding an increase in margins.

The narrowing of skill differentials was associated with a shortage of skilled labor in many spheres of economic activity, although there is no evidence of any causal relationship. The shortage may more properly be explained on the supply side by the impact of the fall in the birthrate in the thirties on the current supply of skill potential⁵ and by restrictive apprenticeship requirements; and on the demand side, by the rapid postwar population growth and industrialization. In fact, the proportion of the 15-19 age group which took up training in the various skilled trades was substantially higher than in the prewar period when a greater premium was offered for skill. It is true, of course, that this contrast may partly be accounted for by the difference in the level of demand. But it is doubtful whether a higher premium for skill in the postwar period would of itself have relieved the shortage of skilled labor. The greater supply of skill rests more on shortening the period of apprenticeship, on more liberal provision of training facilities, and in the longer run, on drawing on the larger numbers of those born in the forties.

It is likely, therefore, that the sharp contraction in differentials between 1947-54 did little more than eliminate a surplus, a payment which was not necessary to induce the required supply of skill, but which was based on the need to maintain a conventional relationship. It is interesting to note that when the Court finally raised the general level of margins in 1954, it did so not to relieve the shortage of skill (since no evidence was brought to relate the shortage to the compressed differentials) but in accordance with the "principles" of wage justice." It did not completely restore the prewar relationship because it observed a tendency operating during and since the war for narrower differentials to prevail.⁶

The foregoing serves to emphasize that the main function of

5. The percentage of the 15-19 age group in the total population fell from 9.13 in 1939 to 6.52 in 1952.

6. The unions demanded a restoration of the 1947 relationship between skilled, semiskilled, and unskilled. But the Court, taking the relationship ruling in 1937 as its basis for adjustment, was able to justify an increase in the margins of the skilled only, arguing that the increase in semiskilled margins in 1947 was not appropriate even though it had been made by mutual approval of employers and unions and confirmed by the Chief Conciliation Commissioner. It is difficult to justify the 1937 basis either economically, since a large measure of unemployment prevailed in the prewar period; or socially, since the 1947 relationship had proved acceptable to both sides. The basic cause of the long strike on the waterfront in January 1956 was the Court's insistence on applying the 1937 relationship to this industry.

relative wage adjustment in Australia has been to distribute the wage bill on the basis of equity considerations, rather than to reallocate the labor force or to provide sufficient incentives for an adequate supply of skill.⁷ Such a policy may be defended on the grounds of industrial expediency since it conforms with conventional practice, especially in a centralized system of wage determination; and on economic grounds, in so far as relative wage movements are not the only or even the most effective means of producing the desired allocation of labor.

The consistency with which the Court has applied income distributional considerations to the relative wage problem is not apparent in its handling of the general level of wages. Since 1931 there has been a marked tendency to use the basic wage as an instrument of general economic policy, such as for reducing unemployment (1931), containing a boom (1937), and more recently, keeping a check on inflation by wage restraint. Nevertheless, in all cases distributional considerations have played a part of varying importance in the Court's decisions. This is implicit in the major principle which the Court has adopted since 1931 to determine changes in the basic wage: the "capacity to pay of the economy," which turns out in practice to be a crude version of keeping the general wage level in line with productivity changes, though not specifically for purposes of pricestability but more to maintain the conventional distribution of income.

The emphasis on equity in the wage-profit relationship is, of course, strongly indicated in the submissions of the unions and the employers' organizations. When one has sifted through all the detailed evidence about the balance of payments, employment, prices, etc., the outstanding point remains the level of profits. The fact that a high level of profits may simply be a reflection of an inflationary pressure is hardly relevant to unions staking a claim for a higher share in these profits before a tribunal on wages. Nor is such an attitude indicative of social irresponsibility. On the contrary the union leaders would be failing in their duty to their members if they did not press a claim to a share of increased profits. The employers' representatives find it, of course, expedient in such circumstances to act as the guardians of the public interest, and to oppose a general wage increase on the grounds of inflation.⁸

What is the Court to do in such a situation? Consider the

7. This is not to deny the need of financial incentives for the supply of skill; but these incentives were more than covered by the existing margins for skill.

8. In depression the parties reverse their positions — the employers pointing to reduced profits as an argument for a reduction in the wage level, and the unions arguing that a wage reduction would aggravate the depression.

1949-50 basic wage claim. Export income was rising sharply as a result of exchange depreciation and the outbreak of the Korean War, stimulating prices and profits in the rest of the economy. The share of wages was falling. (See Table II.) Now an appropriate rise in the real basic wage with subsequent cost-of-living adjustments would clearly produce a substantial increase in prices. But this was the only way in which the "distortion" in the distribution of income could be corrected through *wage* policy.

The Court was sharply divided on what should be done. One side felt that in the public interest the Court should not take any action which would add to inflation. The majority, however, reasoned that the job of the Court was to settle a dispute concerning the workers' "share in prosperity." There was ample evidence that the economy, and particularly the export sector, was prosperous and could bear the burden of higher costs. "If the Court refuses to make any increase in the awarded basic wage because of the inflationary effect of such an increase, then it says in effect that there is no real dispute calling for settlement by the Court or that its proper settlement in the interests of the parties is to refuse the claim *in toto*. I cannot conceive either course having any relation to the realities of the present economic and industrial relations."⁹ The general problem of inflation was a matter more appropriate for government action.

By any economic standard, this was a deplorably clumsy method of redistributing income, particularly as the export boom could not be expected to last indefinitely. By the time export prices were beginning to fall, the wage level was rising faster than ever. Nevertheless, the action of the Court may be defended on several grounds. First, to have ignored the claim for a higher real wage in the interest of price stability when a higher level of national productivity had clearly been achieved (it is true, through an improvement in the terms of trade), would have thrown the Court into disrepute and possibly disuse, with considerable disruption in industrial relations. And after all, whatever its statutory function, any tribunal interposing between unions and employers, must have regard to the impact of its decisions on industrial peace.

An example of a disregard for the realities of the situation in deference to price stability is shown by the decision of the Court in 1953 to discontinue a time honored feature of Australian wage-fixing: the automatic cost-of-living adjustment system. So long as prices remained stable, this decision produced little more than a nominal

9. *The Basic Wage Enquiry, 1949-50*, Print No. A1467, p. 112. See also Isaac, "The Claim for a £10 Basic Wage in Australia," *op. cit.*

protest. But there were immediate signs that this decision would lead to confusion. The state tribunal of Queensland decided not to follow the Court's ruling on this issue. And as a political gesture, the Labor government of Victoria ordered its tribunal to continue with automatic adjustments. But when prices began to rise, the clamor for a return to cost-of-living adjustment grew stronger. The Labor government of New South Wales, submitting to strong pressure, legislated for a restoration of automatic adjustments for those working under state awards. In the three most important states, the tribunals fell out of line with the Court on this issue. And although the tribunal of Western Australia has formally discarded the automatic system, it has made a cost-of-living adjustment in advance of the Court. Nevertheless, the Court has reaffirmed its intention not to return to the automatic system, in two successive judgments (May 1956 and April 1957). In each case, however, it raised the basic wage by 4 per cent to offset partly the rise in the cost of living.

As long as prices rise and there is a lag in the adjustment of wages to prices, this juxtaposition of different wage principles is bound to evoke strong protest and industrial unrest. And it will be interesting to see how long the Court will ignore strongly rooted equity notions of the labor market on the matter of cost-of-living adjustments in favor of price stability, and still survive the challenge to its authority from the state tribunals.¹

A second justification for the Court's action in disregarding the inflationary consequences of its 1950 basic wage increase is that to have ignored the claim for a higher basic wage would not have ended the inflation or stopped wages from rising. Sooner or later, the wage level would have risen, and although not as sharply, certainly more inequitably because those employed in government undertakings and in private companies, which are normally committed to adhere to the awards of tribunals, would have had no increases at all. Clearly, a policy of abstinence would have contributed a little to price stability but quite a lot to industrial unrest.

Finally, a positive attack on inflation could only have come from the use of fiscal, monetary, and other measures rather than from wage policy. The Court has no control of these instruments and, indeed, it has no knowledge of how the government proposes to use

1. With the quickening pace of cost-of-living increases, the (non-Labor) federal government called a conference of the state governments in August 1956, with a view to persuading the states concerned to abandon the automatic system. Only Victoria (now under a non-Labor government) responded by legislating for the discontinuance of this system.

them. It cannot assume that the inflationary pressure and the cause of the dispute — the higher profits — will be tackled by the government, especially when signs of such action are completely absent. In fact there are grounds for believing that the government waited on the Court's judgment before taking any action itself. In addition to its other functions, the Court apparently serves also as a convenient scapegoat for government action and inaction!

It should be obvious that if wage policy is to be used at all for general economic objectives such as countering inflation, eliminating balance-of-payments deficits, etc., it must be formulated as an integral part of economic policy. On its own it is likely to work at cross purposes with the other instruments of policy, especially if the distribution and equity demands of an important section of the economy are not met by measures other than wage changes.

The contribution of wage policy to general economic policy must not be overrated. Even as an integrated part of general economic policy, the role of wage policy is likely to be a minor one. It can only play a negative part in maintaining price stability, for as long as excess demand exists, overaward payments in one form or another will operate to defeat the restraining hand of wage policy. In the allocation of labor, it has been argued that unless there are serious structural problems, it would be more effective to rely on differential profit margins to produce the desired allocation of labor. As a means of promoting recovery from depression, wage policy is of dubious value. And as a means of correcting balance-of-payments disequilibrium by lowering the internal/external cost ratio, it is likely to be severely limited by its impact on the distribution of income. Consider, for example, the use of wage deflation as a method of promoting exports and reducing imports. Apart from the income effects of such a policy, (which could be produced equally by fiscal policy), the effect on the balance of payments of the reduction in wage costs would depend on the elasticity of supply of exports and the elasticity of demand for imports. Because of the nature of Australian exports and imports, it would appear that these elasticities are very small. Consequently, a large reduction in wages and thus a large redistribution of income from wages to profits in the relevant sectors would be necessary to provide the necessary adjustment in foreign trade. The social and political implications of such a policy would be sufficiently important for alternative methods of correcting the disequilibrium in the balance of payments to be used.

Clearly, the burden of general economic policy will fall only to a minor extent on wage policy. And even the income distribution

impact of wage policy, which we have suggested is the major consideration for the Court, is likely to be limited in the long run by the nature of the other policies operating concurrently with it. For example, the squeeze on profits of a higher wage level would be lost if monetary, fiscal, tariff, and other measures responded freely to the higher wages. On the other hand, the creation of a hard market environment² by deliberate use of general economic policy in the face of rising wage rates would be more conducive to a fall in profit margins, or, at any rate, to maintaining real wages at a higher level by inducing productivity increases.

V

A method for formulating and administering national wage policy through the leadership of the Court has developed primarily from the need felt by the general community for a means of settling industrial disputes and promoting industrial peace; and from a desire on the part of the unions and employers for uniformity in wage movements and for a "just" distribution of the national product.

To impose upon the present method of implementing wage policy broader objectives which may involve a departure from accepted notions of equity, is likely to strain the authority of the Court to the breaking point. More than any other instrument of policy, the determination of wages by the government or a statutory authority rests upon its acceptability to unions and employers.³

Furthermore, the composition of the Court, its lengthy legal procedure, and most importantly, its separation from those formulating general economic policy, inevitably restricts the Court to a more limited function. The Dutch⁴ and Norwegian⁵ experiences clearly emphasize the need for an integration, formally or informally,

2. See E. H. Phelps Brown, "Long Term Movements of Real Wages," in J. T. Dunlop (ed.), *The Theory of Wage Determination* (Macmillan, 1957), chap. IV.

3. Cf. Clark Kerr, "Governmental Wage Restraints: Their Limits and Uses in a Mobilized Economy," *American Economic Review*, XLII (May 1952), 384: "The public members on a wage board should be above all else good mediators and the ideal mediator cares mostly about the fact of an acceptable settlement. The economist qua economist is apt to care too much about the content of the settlement, although acceptability is, and by the nature of the situation must be, the ultimate reference. The general public should look to the public members to mediate industrial peace and to the President and Congress, if anywhere, to stabilize the economy."

4. See B. Zoetewij, *op. cit.*

5. J. Inman, "Postwar Wage Policy in Norway," *Bulletin of the Oxford University Institute of Statistics*, Vol. 12 (July-Aug. 1950); M. W. Leiserson, *op. cit.*, p. 139; W. Galenson, *op. cit.*, p. 161.

of wage policy with other instruments of economic policy, if it is to serve, in its limited way, the wider objectives of social and economic policy.

It should be stressed that the need for a national wage policy in Australia does not arise primarily from the belief that wage policy as such is an indispensable part of general economic policy or of economic "planning." Rather, a national wage policy in Australia is an institutional necessity, arising from the development of centralized institutions — labor, employers, and arbitration — and key wage rates — the basic wage and the fitter's margin — which have not only created a means of implementing wage policy on a national scale, but also a continuing need for such a policy. Moreover, on key wage matters there is a decided preference on the part of the central representatives of labor and employers for the arbitral decision of a statutory authority as against the resolution of such matters by free collective bargaining.

Thus the relevant question for Australia is not whether a national wage policy is desirable or not, but whether it should be implemented by an autonomous tribunal dissociated from the major instruments of economic policy, balancing on a tightrope between equity and industrial peace on the one hand, and the wider objectives of national policy on the other. Or whether it behooves the government to grasp the nettle and bring wage policy under its control.

It is likely that the intervention of government directly in wage policy would be acceptable, at any rate to the unions. Government intervention is not only common in many phases of economic life but it is for many an article of faith. There is no strong tradition of free collective bargaining, and, moreover, the highly centralized union and employers' organizations would make a "triangular" bargaining unit feasible.

However, under the present constitutional provisions, the federal government has no powers to fix wages and prices, and consequently, a change in the constitution by referendum or a cession of these powers by the states would be necessary to provide the means for a unified economic policy.⁶ It is unlikely that anything short of a full-blooded socialist government would be bold enough to initiate such a move and be dragged into the political difficulties of fixing wages and profits. It is an interesting commentary on the political volatility of the wage-profit question that while governments freely accept the duty of varying personal incomes through taxation and social

6. In 1945 a referendum proposed by a Labor government for wider industrial powers, including the power to legislate directly on wages, was defeated.

services, they nevertheless shrink from the task of fixing incomes more directly. Instead, this task is delegated to an independent quasi-legislative body which is beyond the popular vote but which is necessarily restricted to a narrow view of the function of wage policy.⁷

This situation would be tolerable if the limitations of such a wage policy were clearly appreciated. Experience shows, however, the ease with which, in the interest of political expediency, the Court becomes saddled with a whole host of jobs most of which it is not competent to perform — maintaining full employment and price stability, adjusting to changes in the terms of trade and the balance of payments, promoting a balanced economic development, and also maintaining industrial peace. When, after a long and patient inquiry the Court has finally pronounced its judgment, fiscal, monetary, and other instruments of policy, limited by the political necessity of maintaining high employment, are then brought into play to accommodate or counteract the effects of the wage changes ordered by the Court. This is a labor standard⁸ with a vengeance.

7. The present (non-Labor) federal government has made its position quite clear recently. "As to wages, (the Government's) view is that they should be determined by independent tribunals, which are the proper bodies to assess the issues, weigh the facts, and apply consistent principles of justice as between wage-earners and employers. It also believes that there is much to be gained from a substantial uniformity throughout Commonwealth and State jurisdictions in the principles and practices of wage determination." The Federal Treasurer's Budget Speech, August 1956.

8. J. R. Hicks, "The Economic Foundations of Wage Policy," *Economic Journal*, LXV (Sept. 1955).

UNIVERSITY OF MELBOURNE

J. E. ISAAC.

MOTOR CARRIER COSTS AND MINIMUM RATE REGULATION

By HOWARD W. NICHOLSON

I. California experience illustrates theoretical and practical difficulties of minimum rate regulation, 139. — II. Description of California's system of controlling minimum rates of intrastate carriers, 140. — III. Analysis of the effects of this policy, 144. — IV. Some fundamental objections to comprehensive systems of minimum rate control, 148.

I

Growing competition in transportation during the twentieth century has exerted downward pressure on the rate structure. This has resulted in increasing attention to problems of the lower limits of rates. Minimum rate regulation has become a topical and important question. Evidence of this is afforded by the current controversy over policy at the national level.¹

Minimum rate regulation raises interesting theoretical as well as practical questions, questions which deserve more detailed examination than they have received. Several recent articles have suggested that comprehensive minimum rate policies be devised which will base rates on either the fully allocated or marginal costs of low cost rail and highway carriers.² But general criticism of such proposals may be raised. Is it feasible or desirable to attempt comprehensive minimum rate control for both rail and highway carriers? If competition forces rates downward, is it not possible and desirable to rely upon the operation of market forces to regulate the lower level of rates?

This paper deals with some of the practical and theoretical problems involved in minimum rate regulation. It includes a study of an ambitious attempt at minimum rate regulation by one of the states. For more than twenty years California has controlled minimum rates charged by intrastate carriers. California's experience demonstrates the fact that fundamental problems confront those

1. Recent recommendations of the Presidential Advisory Committee on Transport Policy and Organization have emphasized the urgency of the practical problem of developing a workable minimum rate structure.

2. See Truman C. Bigham, "Regulation of Minimum Rates in Transportation," this *Journal*, LXI (Feb. 1947), 206-31; also, George W. Wilson, "Effects of Value-of-Service Pricing upon Motor Common Carriers," *Journal of Political Economy*, LXIII (Aug. 1955), 337-44.

who would attempt to frame workable minimum rate policies for highway carriers.

The following discussion contains a description of the essential features of California's minimum rate policy followed by an account of certain problems which have arisen from this type of regulation. The paper concludes with an analysis of the reasons for the types of difficulties which have been encountered, and a discussion of the merits of some alternative policies.

II

California's policy seems to be similar to those advocated by writers who have stressed the merits of fully allocated costs as legal floors to rates. Professor Truman C. Bigham has advocated that minimum rate policies be based upon fully allocated costs with high cost carriers given limited opportunity to meet the rates of their low cost rivals. Recently George W. Wilson has stressed the merits of a minimum rate policy which would authorize a rate structure lying between the average total costs of motor and rail carriers. For types of traffic in which motor carrier costs were lower than rail costs Wilson advocated that motor carriers be authorized to reduce rates to the computed average total costs of the highway carriers to face contract and private carrier competition.³ The California system of controlling minimum rates seems to correspond to the policies mentioned above in that rate floors for regulated carriers tend to be based upon the costs of rail or highway carriers according to which is the low cost producer.

Minimum rate regulation in California was adopted in the 1930's as a means of curbing cutthroat competition between trucking firms and railroads. Statutory changes in 1935, subsequent decisions by the Public Utilities Commission, and supplementary legislation in later years led to the development of a regulatory policy under which for-hire carriers fell into four classes involving two broad categories. The four classes consist of: (1) Common carriers, both rail and highway, which operate between fixed termini and over regular routes. (2) Highway common carriers whose operations are not between fixed termini or over regular routes. (3) Intercity contract carriers. (4) City carriers, both common and contract. Carriers in Class 1 are known as certificated carriers. Classes 2 to 4 are known as permitted carriers. There are important differences in conditions of entry and mechanics of rate regulation for these two broad groups.

3. *Ibid.*

Certificated carriers are regulated under the Public Utilities Act.⁴ A certificate of convenience and necessity is required before commencing operations, and tariffs containing rates must be published, filed, and strictly adhered to. Rates may not be increased without a showing that such changes are justified. Permitted carriers are regulated under the Highway Carrier's Act and the City Carriers Act.⁵ They may begin operations by securing a permit which is issued upon the carrier's securing adequate protection against liability as provided in the Acts. While required to observe as minimum whatever minimum rates may be established by the Commission, they do not file their actual rates and charges with the Commission.

Full competition in meeting the rates of other for-hire carriers is permitted under the California minimum rate policy. In various decisions the Commission has developed the following principles concerning this point. Rail lines and trucks may meet each other's rates.⁶ Highway carriers may not be denied the right to meet competitive rail rates in view of the Highway Carrier's Act, Section 10, even though the rates may not be compensatory.⁷ Rail lines are not permitted to establish rates lower than the minimum rates of the highway carriers without making those rates the applicable minimum rates of the highway carriers.⁸ Authorization for highway common carriers to file rates below existing minima for permitted carriers applies to all shippers and such rates upon publication in the highway common carrier tariff are available for adoption by permitted carriers without further order of the Commission.⁹

To equalize competitive conditions and prevent excessive rate cutting the Commission decided to determine which was the rate-making carrier for various types of shipments.¹ The rate structure includes two broad categories, class rates and commodity rates. The truck was considered by the Commission to be the rate-making carrier for class rates for less-than-truckload and less-than-carload shipments. Rails and vessels and in some cases trucks were held to be rate-making carriers for class and commodity rates for truckload and carload shipments.

4. The Public Utilities Act of California was enacted in 1911 and became effective March 23, 1912. It is now incorporated in the State of California Public Utilities Code, Statutes 1951, Chapter 764.

5. Enacted as the Highway Carriers' Act (Statutes 1935, Chapter 223) and the City Carriers' Act (Statutes 1935, Chapter 312) and subsequently incorporated into the State of California Public Utilities Code in 1951.

6. *Re Petroleum Products*, 40 C.R.C. 221, 258.

7. *Re Canned Goods*, 41 C.R.C. 184, 194.

8. *Minimum Rates*, 44 C.R.C. 108, 118. (P.U.A. Sec. 32d.)

9. *Asbury Transportation Co.*, 49 Cal. P.U.C. 267, 268.

1. *Re Transportation Rates*, 41 C.R.C. 671, 676.

The Commission has developed the following principles concerning the determination of minimum rates: Costs of transport are an important element in establishing minimum rates, for generally rates fixed below the cost level will burden other traffic and result in poverty to carriers engaged in transportation at such rates; and rates fixed much above the cost level will restrict the flow of traffic or divert tonnage from the regulated carriers to proprietary trucks.² It is not the intent to hold up an umbrella or deprive the public of the benefit of reduced rates.³ Minimum rates should not be designed to protect the revenues of all carriers.⁴ Minimum rates should be related closely to particular services.⁵

As a means of implementing this policy the Public Utilities Commission has published tariffs containing minimum legal rates for permitted carriers. In connection with the establishment of these tariffs careful studies have been made by the engineers of the Commission staff to obtain reliable data on motor carrier costs.

The care taken by Commission engineers in determining motor carrier costs is probably unsurpassed by any regulatory agency in the United States. Skill and thoroughness have been shown both in the collection and in the analysis of data. The data collected have included not only information supplied by trucking firms but also the results of time and performance studies made by Commission engineers in the field. The technique used in deriving costs is capable of a high degree of refinement in pinpointing out-of-pocket (or direct) and constant (or indirect) costs of particular types of trucking service. Fully allocated costs for rate-making purposes are determined by applying indirect operating costs on a percentage basis to direct operating costs.⁶ Estimates have been made of costs incurred by efficient motor carriers for a variety of services under widely different types of operating conditions. Cost estimates cover cases ranging from truckload shipments of specific commodities to less-than-truckload shipments of general commodities in operations involving pickup and delivery as well as terminal handling and line haul costs. Studies have been made of the costs of various lengths of haul, sizes of shipments and types of commodities. In addition to considering general

2. *Re Minimum Rates*, 40 C.R.C. 837, 841.

3. *Re Alcoholic Beverage Rates*, 43 C.R.C. 25, 37.

4. *Re Furniture Rates*, 42 C.R.C. 119, 129.

5. *Petition of Truck Owners Association of California et al.* D 50297, C 5432 (Petition 17) (July 20, 1954).

6. Public Utilities Commission, State of California, Transportation Department Research Division, *Report on the Cost of Transporting Property by Motor Vehicle Equipment in the State of California* (Case No. 4808), Dec. 1, 1948.

commodities, special attention has been given to the costs of shipping livestock, used household goods, petroleum and petroleum products in tank cars, property transported in dump trucks, fresh fruits and vegetables, cement, uncrated furniture, and vehicles transported by truckaway service.⁷

For classes of traffic for which the motor truck is regarded as the rate-making carrier, rates charged by all regulated carriers have tended to gravitate to the level prescribed as minimum rates for permitted carriers in the Commission's tariffs. Although the Commission has generally taken the position that rail rates should determine the legal minimum for carload and truckload shipments, so little information has been available about rail costs that there has been a tendency to base class rail rates for carload shipments on truck rates.⁸ To the extent that this has been done, truck rates have become established as minima for rail carload as well as for less-than-carload and less-than-truckload shipments that move on class rates.

The bulk of rail carload traffic, however, moves under commodity rates. These are generally lower than class rates and serve as legal rate floors for both carload and truckload shipments.

It is the policy of the Commission that requests for exemption from minimum rates must be decided in accordance with the facts in each case.⁹ The general rule is that exemptions are granted where it is shown that minimum rates which have been established are inappropriate or unsuitable for particular traffic, and where timely modification of the rates is impracticable.¹ The exemption of carriers by name is not favored, but has sometimes been permitted as a matter of expediency.²

7. Fifteen separate tariffs naming minimum rates, rules and regulations for the transportation of property between points in California are now published by the California Public Utilities Commission. These include City Carriers' Tariffs Numbers 1-A, 2-A, 7; Highway Carriers' Tariffs Numbers 1-A, 9; and Minimum Rate Tariffs 2, 3, 4-A, 5, 6, 7, 8, 10, 11-A, 12. Minimum Rate Tariffs 3, 4-A, 6, 7, 8, 10, 11-A and 12 contain rates for livestock, used household goods and office fixtures, petroleum products, property transported in dump trucks, fresh fruits and vegetables, cement, uncrated new furniture and truckaway service respectively. City Carriers' Tariffs 1-A, 2-A, and 7 contain rates for San Francisco City and County, Metropolitan Oakland, and Metropolitan San Diego. Highway Carriers' Tariffs No. 1-A and 9 contain rates for Metropolitan Oakland and San Diego, and Minimum Rate Tariff Number 5 contains rates for the Los Angeles Drayage Area. The other Minimum Rate Tariffs contain rates for state-wide shipments. A distance table containing constructive highway mileages is used in connection with these tariffs.

8. See the testimony of Mr. Meinhold, representative of the railroads at the *Hearings before the Public Utilities Commission of California, Nov. 26, 1951 in San Francisco*. Reporters Transcript. Case 4808 Commission, Volume 116.

9. *H. G. Arnen and F. Helen Arnen (Metro Parcel Service)*, 47 Cal. P.U.C.

1. *Ibid.* 2. *Ibid.*

III

It would appear that California has developed a policy whereby legal minimum rates are related to the fully allocated costs of low cost carriers without placing any restriction on price competition between carriers as long as their rates do not fall below legal minima. More detailed examination reveals, however, that there are important discrepancies between fully allocated costs and minimum rates.

Discrepancies between fully allocated costs and the minimum rates which highway carriers are permitted to charge for particular services are neither uncommon nor unimportant.³ On the contrary, they are often the intended and sometimes the inevitable results of the state's policy.

California shippers have argued before the Commission that cost differences between broad classes of carriers have not been properly reflected in the minimum rate structure.⁴ In 1951 shippers complained that highway contract carriers enjoyed lower operating costs than highway common carriers, but that the published minimum rates for permitted carriers were designed to return "average" costs for all carriers including the common carriers (whose rates gravitate to the same level as those prescribed for permitted carriers). As a result, it was argued, the published minimum rates must be excessive for the low cost carriers.⁵

The Commission disagreed with these arguments holding that the shippers did not appreciate the degree of refinement of the Commission's cost finding techniques.⁶ Despite the extent of refinement of the cost analysis, however, cost estimates made for the purpose of minimum rate regulation for large numbers of carriers must represent an average figure for operators operating under the assumed conditions. Costs of individual carriers or even of groups of carriers may vary more or less consistently above or below the Commission's estimates.

Apart from inflexibility in the application of the cost formula a second reason why fully allocated motor carrier costs and legal minimum rates may not coincide is the manner in which published minimum rates are derived from the cost estimates.

Minimum rate tariffs published by the California Public Utilities Commission are not simple extensions of cost formulas. In a discussion of the factors considered in rate making J. A. McCunniff,

3. There is so little information available about rail costs that not much can be said about the relationship between rail rates and fully allocated rail costs.

4. See Reporters Transcript. Case 4808 Commission Vol. 116, *op. cit.*

5. *Ibid.*

6. Public Utilities Commission Decision 46.

Senior Rate Analyst for the Commission, has emphasized that under the law the Commission is enjoined to take into account not only costs of producing the service but also the value of the commodity transported and the value of the facility reasonably necessary to perform such transportation service.⁷ Furthermore, this mandate does not preclude other rate-making factors. Judgment in arriving at published minimum rates must also be exercised in consideration of what the traffic will bear, in assigning commodities to classes in the selection of a classification, in establishing relationships between classes, and in establishing weight brackets.⁸

The use of judgment in dealing with the above mentioned factors involves a weighing of value of service considerations as well as an analysis of cost factors in arriving at the published legal minimum rates. In this way the procedure used in translating costs into rates changes the rate-cost relationship by explicitly introducing demand factors into rate determination.⁹ Thus, although the Public Utilities Commission has taken great care in collecting and analyzing data relating to the costs of truck shipments for use as a basis for minimum rate tariffs, legal rates for shipments of specific commodities by truck not infrequently may differ from estimates of the fully allocated costs of producing the service in question.

Inflexibility in the application of cost formulas and the use of demand in addition to cost factors in the determination of rates explain the numerous discrepancies which exist between legal minimum rates and fully allocated costs of producing particular services. In addition, however, there is evidence that the general level of minimum rates established in California has become too high.

Minimum rate policies result in levels which may be characterized as too high when rates tend to be higher than necessary to cover the average total unit costs of efficient carriers capable of providing the quality of service desired by shippers. Normally rates are considered to be too high from the standpoint of consumers of transport service. From the standpoint of the carrier high rates are not objectionable assuming that each carrier is able to sell the volume of output he wishes to produce at these rates. But if this assumption

7. J. A. McCunniff, *Study of Economic Considerations in the Establishment of Minimum Transportation Rates and Related Problems*, Vol. 1 (Public Utilities Commission of California, San Francisco, California, 1949), p. 55.

8. *Ibid.*, Vols. I and II. See especially Part 2, Vol. I and Part 3, Vol. II.

9. The fact that the evaluation of demand factors is made by a regulatory agency rather than by individual carriers or bureaus representing carriers may produce results which are not consistent with the interest of either shippers or carriers.

is not valid, rates may be too high from the carrier's standpoint — i.e., if they price the carrier out of the market or otherwise curtail his sales so that his profits are reduced.

Rates may be too low from the producer's standpoint if they fail to cover average total unit costs of services which shippers desire. Normally rates are never too low from the consumer's standpoint assuming that the desired quality of service is not impaired by inability of producers to obtain rates sufficiently high to cover costs.¹

California originally adopted a comprehensive minimum rate policy because the going rates of carriers threatened to become too low. Pressure for minimum rate regulation was exerted by producer interests for obvious reasons, but evidence existed that consumer interests might be adversely affected by rates which were so low that the quality of transportation service would deteriorate.²

Unhappily the minimum rate policy adopted as a remedy seems to have resulted in making rates for regulated highway carriers too high. This seems to be the result of the operation of forces which the Public Utilities Commission has not been able to control.³

Costs of for-hire highway carriers apparently are not independent of the minimum rates they are permitted to charge. Minimum rate regulation together with unrestricted competition between for-hire and nonregulated carriers seems to have resulted in an unstable situation in which upward pressure is exerted on for-hire carrier costs. This pressure is aggravated by methods of determining minimum rates which produce rates for specific services that are higher than the true minimum cost of producing that service.⁴

1. In view of the importance of quality characteristics in for-hire transportation service the possibility that cutthroat competition may sacrifice dependability and other important qualities of service cannot be entirely discounted. This is particularly true of common carriers whose obligations to shippers under the law are quite extensive.

2. P.U.C. Decision No. 42646 of March 22, 1949 in Case No. 4823 for a review and discussion of the antecedents to present statutes under which for-hire carriers are regulated.

3. Minimum rates established for permitted carriers have come to be regarded as going rates for certificated as well as permitted carriers. Also, minimum rates published in Commission tariffs include a "normal profit" as an element of costs.

4. Reasons why minimum rates for shipping specific commodities between designated points may differ from the fully allocated or true minimum costs of providing particular shipping services have been referred to above. Also, discrepancies between rates and costs may arise as a result of general rate increases made to adjust the level of rates for increases in costs during periods of inflation. Across-the-board rate increases provide a relatively simple method of adjusting the rate structure, but they often tend to increase the number of cases in which minimum rates exceed the fully allocated costs of providing particular services.

The way in which this upward pressure is exerted on carrier costs may be described in the following manner: Unregulated carriers tend to be attracted into the shipment of commodities for which legal minimum rates are higher than costs. Competition by unregulated carriers thus tends to erode the relatively profitable traffic of regulated carriers, and profits of for-hire carriers tend to decline. This may be accompanied by a decline in traffic volume and load factors for regulated carriers which tends to increase average costs. Both the fall in profits and rising costs tend to lead to an increase in the level of legal minimum rates. Higher minimum rates tend to attract more permitted carriers (whose entry in effect is not subject to restriction). It also results in an increased tendency for the erosion of profitable traffic by private carriers. This tends further to reduce profits of for-hire carriers, increase costs, etc.

In the manner described above, competition coupled with minimum rate regulation creates rates which are "too high" from both a shipper and carrier standpoint. Not only do rates tend to rise above the costs of shipment by efficient, fully employed for-hire carriers, but also the increase in rates tends to reduce the sales of for-hire carriers and to cut their profits.

The Public Utilities Commission of California has recognized the existence of these difficulties and has attempted to deal with these problems. The Commission has emphasized the principle that minimum rates should be related closely to particular services. (See page 142 above.)⁵ Efforts have been made to refine cost analysis and to increase the number and quality of minimum rate tariffs.⁶ Minimum rates have been suspended in cases where they were causing a diversion of traffic to proprietary competition.⁷ Also, the Commission has advocated (although unsuccessfully) that effective means of limiting the entry of permitted carriers be authorized by the legislature.⁸

Yet despite these efforts rates still tend to be "too high." Perhaps the best informed criticism of the results of California's experience with minimum rate control has been made by a former member of the

The California Public Utilities Commission has often resorted to across-the-board rate increases as an interim means of raising legal minimum rates to keep pace with rising costs.

5. *Petition of Truck Owners Association of California et al.* D 50297, C 5432 (Petition 17) (July 20, 1954).

6. Cf. pages 142 and 143 above.

7. An example of this is given in P.U.C. Decision No. 45337, Application No. 31809, Feb. 6, 1951, in which the Commission authorized less-than-minimum rates for certain canned goods shipments.

8. P.U.C. Decision No. 41470, Case 4823, April 13, 1948. 48 Cal. P.U.C. 52.

California Public Utilities Commission, now a member of the Interstate Commerce Commission. Mr. Howard Freas has observed that California's policy tended to eliminate price cutting, but the result was a rate structure higher than it would have been under free competition. What were theoretically minimum rates became going rates if not maximum rates. If rates were set on a true minimum level, they were inadequate as going rates, and if proper as going rates they exceeded the minimum level which the statute contemplated.⁹

IV

The conclusion seems warranted that minimum rate control in California has not been successful. But it is not immediately evident why the policy has proved to be inadequate. The reasons deserve careful examination for they involve fundamental problems of minimum rate control.

Analysis of the reasons for the failure of California's minimum rate policy should begin with the recognition of one basic point. The cost characteristics of highway transport rule out the feasibility of basing minimum rates on the fully allocated costs of producing particular services.

This statement requires explanation for it seems to disregard important facts. Students of the industry have concluded that highway transport costs are almost 100 per cent variable. Furthermore, it is well known that the cost of providing particular types of highway transport service are relatively easy to determine. Indeed, one of the most striking features of highway transportation is the extent to which specialization has occurred in this field to minimize the costs of providing particular services. Trucks are typically custom built in order to produce more efficient transport service for specific commodities. So-called common carriers in highway transportation specialize in handling particular commodities. Both the high percentage of variable costs in trucking and the degree to which the costs of providing particular services can be distinguished would make it appear that it is not impractical to estimate fully allocated costs of producing particular services and to establish these as realistic minimum rates.

But for purposes of minimum rate regulation one cannot ignore the fact that highway transport service is produced under conditions of joint cost. Back hauls and time jointness characterize the supply of highway transportation and create conditions which in many cases

9. Howard Freas, "Traffic Rates, Charges, Tariffs and Costs," *Interstate Commerce Commission Practitioners Journal*, June 1955, pp. 933-34.

make it illogical and impractical to arrive at definitive estimates of fully allocated costs. It is well known that under competitive conditions the only sound economic rule for pricing services subject to joint costs is to try to insure that total revenue from the sale of all services covers total cost. In order to accomplish this, demand as well as cost factors must be taken into account in order to determine the correct price.

To the extent that highway transport services are produced under conditions of joint cost it is therefore impossible to establish correct minimum rates by pure cost analysis. If minimum rates disregard demand factors, regulated carriers may be prevented from fully exploiting potential business. Threatened by unregulated competition, regulated carriers may lose business which in the absence of rate regulation they might readily secure.

Since joint cost characterizes the production of highway transport service, demand as well as cost should be taken into consideration in determining minimum rates. But what bearing does this have on the difficulties experienced with California's minimum rate policy? As we have noted, demand as well as cost factors are taken into account in the determination of minimum rate tariffs. What is the basic flaw in California's system?

The answer to this question is that California's policy lacks sufficient flexibility to provide a workable minimum rate structure. Despite the fact that careful study is made of carriers' costs, that minimum rates are derived from cost data only after consideration of demand factors, and that provision is made for modification of rates which cause regulated carriers to lose traffic, there is not sufficient flexibility in the policy to eliminate serious disadvantages to carriers and consumers of for-hire carrier service.

Lack of sufficient flexibility is not due to administrative failure. The problem is deeper than that. It is one thing for a firm producing services subject to joint costs to arrive at correct prices — prices which reflect both cost and demand factors. It is quite a different thing for a government agency to attempt to carry out this function for thousands of firms subject to its jurisdiction — firms confronted with varying demand conditions in numerous special markets.

Minimum rates for specific services established by a regulatory agency for highway carriers will almost invariably differ from prices which would be established by individual carriers free to set rates according to market conditions. The existence of unregulated competition coupled with the degree to which costs of providing particular types of highway transport may be distinguished and reduced by

specialized carriers will increase the practical significance of errors made by regulatory bodies in setting minimum rates and will penalize carriers subject to rate controls. The only system that will insure sufficient flexibility to enable producers of services subject to joint costs to establish correct prices at all times would involve the abandonment of minimum rate control.

This analysis indicates that policies designed to control minimum rates for highway carriers prove to be unsatisfactory, first, because no logical basis exists for basing rates solely on costs in view of the significance of joint costs and, second, because of the inability of a regulatory agency to establish correct rates by taking demand factors into account. This seems to lead to the conclusion that no regulation of minimum rates should be permitted.

Unfortunately, there are grounds for believing that serious difficulties may arise if minimum rate controls are completely abandoned. The prevalence of joint costs creates a danger that unrestricted competition may cause rates to tend to fall below remunerative levels when change occurs in highway travel.

The problem arises because of the importance of back hauls. Consider a carrier that ships a commodity between points *A* and *B*. Whenever this transportation service is produced, as a by-product there arises a supply of transportation service between *B* and *A* on the return haul. The rate which a carrier is willing to accept for hauling commodities from *B* to *A* may be very low for it will be profitable as long as it more than covers the one way out-of-pocket expenses which are separable from the variable joint expenses incurred in the round-trip movement of the equipment. Thus for a carrier based at *A* who accepts shipments to *B* at a remunerative rate, there is incentive to transport shipments from *B* to *A* at a rate which may be lower than one-half of the variable costs of the round trip between *A* and *B*. If there is only one carrier there is, of course, no likelihood that the sum of rates for services produced between points *A* and *B* will fall below the variable costs of producing the service. But assume that a second carrier located at *B* originates traffic to *A*. Just as the first carrier on the back haul from *B* to *A* might find it profitable to charge rates lower than one-half of the variable costs of the round trip from *A* to *B* so a carrier located at and originating traffic at *B* may be tempted to charge rates on his back haul which are lower than one-half of the variable costs of the round trip from *B* to *A*.

In short, the competitive floor to rates from *B* to *A* consists of the relatively small added expenses incurred by the back-haul carrier

for the shipment of a return cargo, while the floor to rates from *A* to *B* consists of the relatively small added expenses incurred by the back-haul carrier in this direction. The sum of these rates may be considerably below the out-of-pocket expenses which one carrier would incur in a round trip operation between these two points. Thus the joint cost nature of the highway transport service which is produced by the back haul may create pressure for cutting rates below variable costs for shipments between any two points.

The analysis given above explains how rates may be forced below the variable costs of highway carriers between given points. It is true, of course, that round trip rates which total less than out-of-pocket costs will soon force carriers to curtail operations and to raise rates. It will be noted that our original assumption was that the carrier which accepted a low back-haul rate had received a remunerative rate for the original shipment. Indeed there would be no incentive for a carrier to begin a round trip journey if he could not expect a return sufficiently high to cover variable costs.

But while a natural corrective for unduly low rates exists, the danger of a chronic tendency toward cutthroat competition in highway transport cannot be completely discounted. The creation of the back haul as a by-product of an original haul means that whenever profitable traffic is generated between given points surplus capacity develops in the opposite direction. This surplus capacity overhangs the market and tends to depress rates to unremunerative levels. It is not easy to generalize about the seriousness of this factor. Flexibility in the choice of routes and the ease with which firms may enter the industry may aggravate this tendency to force rates too low. Previous experience with cutthroat competition in highway transport suggests that unduly low rates may become a serious problem with resulting disadvantages to consumers as well as to carriers.

If this is so, is there any solution to the problem of minimum rate policy?¹ May we not conclude that for all of its defects the type of system developed by California is acceptable? It is, of course, easy to criticize policy, but is there a more satisfactory alternative?

On general grounds the basic principle of minimum rate regulation in California cannot be defended. California's experience illus-

1. To the extent that competition forces rates to low levels an obvious method of dealing with the problem is to attempt to control competition. Our analysis is confined to the problem of minimum rate regulation abstracting from the policy of attempting to control competition through regulating the number of carriers. In California, of course, competition from unregulated carriers has characterized the industry and little effective control has been exercised over the entry of permitted carriers.

trates the fact that comprehensive systems of minimum rate control tend to raise rates and put the burden of proof on those who claim that minimum rates are too high. Despite safeguards designed to prevent cost analysis from casting rates into an unreasonable mold, despite provisions designed to lower rates when they prove to be excessive, the results of comprehensive minimum rate laws are likely to be disappointing. In view of the cost conditions in highway transportation mistakes in establishing minimum rates cannot be avoided, and the basic drawback of California's policy is that these mistakes are not easy to rectify, resulting in disadvantages for both regulated carriers and consumers.

A more satisfactory alternative would be to place the burden of proof on those who would claim that rates are too low. This means that policies which attempt to base minimum rates on general formulas should be replaced by a policy which permits rates to seek their natural level.

The danger that rates may tend to settle at levels which are not remunerative may be recognized by the provision that more or less temporary floors to rates may be established in those cases where serious problems develop. Periodic review of these cases could be made which would result in discontinuance of rate controls unless a clear need for their continuance could be shown. Such a policy would not eliminate all difficulties. But it would seem to be more consistent both with the general interests of consumers and regulated carriers that the burden of proof be placed on those who claim that rates are too low. California's experience is a lesson that the protection that should be provided against the eventuality of excessively low rates should not be of such a type that it produces a constant tendency for minimum rates to become too high.²

2. Commissioner Freas has advocated a policy similar to that which is recommended in this paragraph. See Howard Freas, *op. cit.*, pp. 933-34.

HOWARD W. NICHOLSON.

UNIVERSITY OF VIRGINIA

RECENT PUBLICATIONS

- ALDERFER, E. B. AND MICHL, H. E. *Economics of American Industry*, Third Edition. New York: McGraw-Hill Book Company. 1957. pp. 710. \$7.00.
- BAKER, ELIZABETH FAULKNER. *Printers and Technology*. New York: Columbia University Press. 1957. pp. 545. \$7.00.
- BANERJI, ANJAN KUMAR. *Story of the Swelling Numbers in India*. Calcutta, India: Benson's. 1956. pp. 48.
- BANK FOR INTERNATIONAL SETTLEMENTS. *Foreign Exchange Regulations in Great Britain*. Supplement 14. Basle, Switzerland. 1957.
- BASU, S. K. *A Survey of Contemporary Banking Trends*. Calcutta, India: The Book Exchange. 1957. pp. 574.
- BEER, SAMUEL H. *Treasury Control*. New York: Oxford University Press. 1957. pp. 139. \$2.40.
- BERGFELD, ALBERT J., EARLEY, JAMES S., KNOBLOCH, WILLIAM R. *Pricing for Profit and Growth*. New York: McGraw-Hill Book Company. 1957. pp. 188.
- BRADLEY, JOSEPH F. AND WHERRY, RALPH H. *Personal and Family Finance*. New York: Rinehart & Company, Inc. 1957. pp. 565. \$6.75.
- CHAMBERLIN, EDWARD HASTINGS. *Towards a More General Theory of Value*. New York: Oxford University Press. 1957. pp. 326. \$5.00.
- CHOW, G. C. *Demand for Automobiles in the United States*. Amsterdam, Holland: North-Holland Publishing Company. 1957. pp. 110.
- CLAWSON, MARION; HELD, BURNELL. *The Federal Lands: Their Use and Management*. Baltimore 18, Maryland: The Johns Hopkins Press. 1957. pp. 501. \$8.50.
- COLE, ARTHUR H. *The Historical Development of Economic and Business Literature*. Boston, Massachusetts: Harvard Graduate School of Business Administration. 1957. pp. 56.
- COMMITTEE ON ECONOMIC POLICY, THE. *The Mechanics of Inflation*. Washington, D. C.: United States Chamber of Commerce. 1957. pp. 70. \$1.00.
- CREIGHTON, DONALD. *Harold Adams Innis, Portrait of a Scholar*. Toronto, Canada: University of Toronto Press. 1957. pp. 146. \$3.00.
- CROSSER, PAUL K. *Economic Fictions*. New York: Philosophical Library. 1957. pp. 322. \$4.75.
- CROWTHER, GEOFFREY. *The Morality of Discrimination in International Trade*. Berkeley, California: University of California. 1957. pp. 29.
- DEHEM, ROGER. *Eléments de Science Économique*. Louvain, Belgium: Éditions Nauwelaerts. 1957. pp. 218.
- DEPARTMENT OF TRADE, MELBOURNE. *United States Investment in Australian Manufacturing Industry*. Melbourne, Australia. 1957. pp. 66.
- DREHER, CARL. *Automation: What it is, How it Works, Who can Use it*. Scranton, Pennsylvania: W. W. Norton & Co. 1957. pp. 128. \$2.95.
- EHRMANN, HENRY W. *Organized Business in France*. Princeton, New Jersey: Princeton University Press. 1957. pp. 514. \$7.50.
- FITZSIMMONS, THOMAS (EDITOR). *Cambodia*. New Haven, Connecticut: HRAF Press. 1957. pp. 350.
- GORRES-GESELLSCHAFT (EDITOR). *Staatslexikon Recht Wirtschaft Gesellschaft*. Freiburg, Germany: Verlag Herder Freiburg. 1957. pp. 1246.
- HABER, WILLIAM AND CARROLL JOHN; KAHN, MARK; PECK, MERTON. *Maintenance of Way Employment on U. S. Railroads*. Detroit, Michigan: Brotherhood of Maintenance of Way Employees. 1957. pp. 237.
- HAILEY, LORD. *An African Survey Revised 1956*. New York: Oxford University Press. October 1957. pp. 1676. \$16.80.

- HOSKINS, W. G. *The Midland Peasant*. New York: St. Martin's Press, Inc. 1957. pp. 322. \$6.75.
- HUNTER, HOLLAND. *Soviet Transportation Policy*. Cambridge, Massachusetts: Harvard University Press. 1957. pp. 416. \$8.50.
- JAMIESON, STUART. *Industrial Relations in Canada*. Ithaca, New York: Cornell University Press. 1957. pp. 144. \$1.75.
- JEANNENEY, JEAN-MARCEL. *Tableaux Statistiques Relatifs A L'Economie Francaise Et L'Economie Mondiale*. Paris: Librairie Armand Colin. 1957. pp. 201.
- JOINT ECONOMIC COMMITTEE. *Federal Expenditure Policy for Economic Growth and Stability*. Washington, D. C.: Government Printing Office. 1957. pp. 1201. \$3.25.
- KATONA, GEORGE WITH LAUTERBACH, ALBERT AND STEINKAMP, STANLEY. *Business Looks at Banks*. Ann Arbor, Michigan: University of Michigan Press. 1957. pp. 184. \$5.00.
- KERR, CLARK. *Productivity and Labor Relations*. Berkeley, California: Institute of Industrial Relations. 1957. pp. 35. 20 cents.
- KOOPMANS, TJALLING C. *The State of Economic Science*. New York: McGraw-Hill Book Company. 1957. pp. 231. \$6.50.
- LUCE, R. DUNCAN AND RAIFFA, HOWARD. *Games and Decisions: Introduction and Critical Survey*. New York: John Wiley & Sons, Inc. 1957. pp. 509.
- LUDWIG, MARIO. *Internationale Rohstoffpolitik*. Zurich, Switzerland: Polygraphischer Verlag. 1957. pp. 128.
- LUNT, DUDLEY C. *The Farmers Bank, 1807-1957*. Dover, Delaware: Farmers Bank of the State of Delaware. 1957. pp. 308. \$10.00.
- McNAIR, MALCOLM P., BROWN, MILTON P., LEIGHTON, DAVID S. R., ENGLAND, WILBUR B. *Problems in Marketing*. New York: McGraw-Hill Book Company. 1957. pp. 740. \$7.00.
- MIERNYK, WILLIAM H. AND HOROWITZ, MORRIS A. *Salaries and Earnings of Engineering Teachers 1956*. Boston, Massachusetts: Bureau of Business and Economic Research, Northeastern University. 1957. pp. 19. 25 cents.
- MORICE, Janine. *La Demande D'Automobiles En France*. Paris: Librairie Armand Colin. 1957. pp. 218.
- MOSER, C. A. *The Measurement of Levels of Living with Special Reference to Jamaica*. London: Colonial Office. 1957. pp. 106.
- NEF, JOHN U. *Industry and Government in France and England, 1540-1640*. Ithaca, New York: Cornell University Press. 1957. pp. 162. \$1.45.
- NOONAN, JOHN T. JR. *The Scholastic Analysis of Usury*. Cambridge, Massachusetts: Harvard University Press. 1957. pp. 432. \$9.00.
- PARKER, H. M. D. *Manpower, A Study of War-time Policy and Administration*. New York: British Information Services. 1957. pp. 535. \$7.20.
- PATTERSON, GARDNER AND FURNISS, EDGAR S. JR. *NATO, a Critical Appraisal*. Princeton, New Jersey: Princeton University Conference on NATO. 1957. pp. 107.
- PHELPS, CLYDE WILLIAM. *Accounts Receivable Financing as a Method of Business Finance*. Baltimore, Maryland. Commercial Credit Company. 1957. pp. 65.
- RATCLIFF, RICHARD U., RATHBUN, DANIEL B., HONNOLD, JUNIA H. *Residential Finance, 1950*. New York: John Wiley & Sons. 1957. pp. 180. \$6.00.
- RAYMOND, ELLSWORTH. *Soviet Economic Progress*. New York. Rinehart & Company, Inc. September. pp. 56. 75 cents.
- ROSS, RALPH AND VAN DEN HAGG, ERNEST. *The Fabric of Society*. New York: Harcourt, Brace and Company. 1957. pp. 777. \$7.50.
- ROYAL COMMISSION ON AGRICULTURE AND RURAL LIFE. 14. *A Program of Improvement*. Saskatchewan, Canada: Government of Saskatchewan. 1957. pp. 149.
- . 11. *Farm Electrification*. Regina, Saskatchewan. 1957. pp. 100. \$1.00

- SAULNIER, R. J., HALCROW, HAROLD G., AND JACOBY, NEIL H. *Federal Lending: Its Growth and Impact*. New York: National Bureau of Economic Research. 1957. pp. 48. \$1.00.
- SHANNON, LYLE W. *Underdeveloped Areas*. New York: Harper & Brothers. 1957. pp. 496. \$6.50.
- SHISKIN, JULIUS. *Electronic Computers and Business Indicators (Occasional Paper 57)*. New York: National Bureau of Economic Research. 1957. pp. 269. \$1.00.
- SPULBER, NICOLAS. *The Economics of Communist Eastern Europe*. New York: John Wiley & Sons, Inc. September 1957. pp. 525. \$12.50.
- SZCZEPANIK, EDWARD F. *The Cost of Living in Hong Kong*. New York: Oxford University Press. 1957. pp. 31. \$1.20.
- TOSTLEBE, ALVIN S. *Capital in Agriculture — Its Formation and Financing Since 1870*. Princeton, New Jersey: Princeton University Press. 1957. pp. 232. \$6.00.
- TROY, LEO. *Distribution of Union Membership among the States 1939 and 1953*. New York: National Bureau of Economic Research. 1957. pp. 32. 75 cents.
- WALKER, GILBERT. *Economic Planning and Programme and Control in Great Britain*. New York: The Macmillan Company. 1957. pp. 175. \$3.75.
- WILLIAMS, ERNEST W. JR. *The Regulation of Rail-Motor Rate Competition*. New York: Harper & Brothers. 1957. pp. 247. \$4.50.
- WINSTON, AMBROSE PARÉ. *Judicial Economics*. Austin, Texas: Ambrose Paré Winston. 1957. pp. 175. \$1.00.
- WRIGHT, DAVID McCORD. *Post-War West German and United Kingdom Recovery*. Washington, D. C.: American Enterprise Association, Inc. 1957. pp. 27.
- ZANIEWSKI, ROMUALD. *L'Origine du Proletariat romain et contemporain*. Louvain, Belgium: Editions Nauwelaerts. 1957. pp. 400. 350 frs b.

ECONOMIE APPLIQUÉE

Tome X

SEPTEMBRE-DÉCEMBRE, 1967

No. 4

PROFIT ET CROISSANCE**

- M. FANNO Profit et Expansion.
J. ROBINSON La théorie de la Répartition.
M. R. FISCHER L'épargne et les profits des entreprises dans l'hypothèse du "revenu permanent."
R. L. MARRIS Le taux de croissance et le degré d'oligopole.
N. SPULBER Taux de profit et croissance dans les économies centralisées et planifiées de type soviétique.
J. L. FYOT Les zones de crise aux Etats-Unis — Problèmes de politique économique régionale.
M. F. CAPET Marges bénéficiaires et formation des prix en A.O.F.
P. ANGOUVENT Préparons l'an 2000.

Economie Appliquée et Cahiers de l'Institut de Science Economique Appliquée
Directeur: François Perroux Secrétaire Général: H. Bonjour
(I.S.E.A., 35 Boulevard des Capucines, Paris 2ème)

INDIAN JOURNAL OF ECONOMICS

(Established in 1916)

Editor: DR. OM PRAKASH

VOL. XXXVIII

JULY 1957

No. 148

Special Number on International Economic Problems

ARTICLES

Symptoms of Economic Development, by *Om Prakash*; An International Economic Policy, by *J. Tinbergen*; Some Reflections on the Concept of "Disguised Unemployment," by *Jacob Viner*; Defects in the Concept of Regionalism to Solve Trading Problems, by *Gottfried Haberler*; Partial — Vs. General — Equilibrium in International Trade, by *C. P. Kindleberger*; Discriminatory Tariff Reduction: A Marshallian Analysis, by *Harry G. Johnson*; Future of World Trade, by *S. L. Parmar*; Overseas Borrowing: The New Model, by *H. W. Arndt*; Money and Banking Problems in the British Colonies, by *A. Hasib*; Some Aspects of Fiscal Policy in Underdeveloped Economies, by *B. K. Singh*; The Problem of Factor Proportions in Underdeveloped Economies, by *A. D. Sharma*; Revealed Preference for Indian Jute, by *George K. Chacko*.

NOTES AND MEMORANDA

Activity, Analysis and the Theory of Economic Equilibrium, by *J. K. Mehta*; Financial Aspects of the Constitution of Pakistan, by *M. S. Bhatia*; A Note on International Economic Relations, by *Mahesh Chand*.

ECONOMIC LITERATURE

Annual subscription (four issues) Rs. 17, £0-26-0 or \$3.80. This Special Number costs Rs. 6.50, £0-10-0 or \$1.50 up to December 31, 1958; and Rs. 8, £0-12-6 or \$1.80 thereafter. Write to the Managing Editor, Indian Journal of Economics, University of Allahabad, Allahabad-2 (India).

The Review of Economics and Statistics

Published by Harvard University

Editor: Seymour E. Harris

Associate Editors: Abram Bergson, Edward H. Chamberlin, John T. Dunlop, John Kenneth Galbraith,* Alexander Gerschenkron, Gottfried Haberler,* Alvin H. Hansen,* Wassily W. Leontief,* John Lintner,* Edward S. Mason, Sumner H. Slichter, Arthur Smithies, John H. Williams.

*Members of Managing Board

Vol. XL

February 1958

Number 1

- Income, Assets and the Demand for Money H. F. Lydall
Monetary Theories of the Rate of Interest: A Dynamic Analysis Warren L. Smith
Employment and Output in Banking 1919-1955 Richard E. Speagle and Ernest Kohn
The Contraction of 1953-54 Bert G. Hickman
Bankers and Subsidies Alvin H. Hansen
Empirical Research and the Rate of Interest Alan Spiro
Accounting for Cost Control in the Soviet Economy Robert W. Campbell
The Rate of Interest on Government Foreign Lending Stephen H. Axilrod
Debt Repayment and Cyclical Fluctuation Harry T. Oshima
NOTES AND BOOK REVIEWS

The Review of Economics and Statistics is published quarterly. Annual subscription price: \$6.00
Order from the Harvard University Press, Cambridge 38, Massachusetts.

THE ECONOMIC JOURNAL

The Quarterly Journal of the Royal Economic Society
DECEMBER 1957

Vol. LXVII

No. 268

I. Articles

Is a Wages Policy Desirable and Practicable?

A Model of Economic Growth

Money, Prices and the Theory of Interest Determination

Expenditure Taxes, Imports and Gross Domestic Product at Market Prices

Depreciation, Replacement and Growth — And Fluctuations

Odds, Possibility and Plausibility in Shackle's Theory of Decision: A Discussion

British-Canadian Manufacturing Productivity

Clive Bell on Keynes (Review Article)

II. Reviews

III. Notes and Memoranda

IV. Recent Periodicals and New Books

D. T. Jack

N. Kaldor

G. Horwich

H. Burton

E. D. Domar

G. Gould and G. L. S. Shackle

J. B. Heath

R. F. Harrod

Macmillan and Co., Ltd., London, W.C. 2

Application for Fellowship to

Royal Economic Society, 21 Bentinck Street, London, W. 1

Annual Subscription £1 10s.

Single Copy Ten Shillings

REVUE DE SCIENCE FINANCIÈRE

THE FRENCH REVIEW OF PUBLIC FINANCE

Number I. January–March, 1958

CONTENTS

Some Current Issues in the Taxation of Personal Income in the United States

Money Flows and Economic Tabulation

The Control of Industrial Location in Great Britain: New Towns

The Solidarity of Tax-payers

Tax Capitalisation and Investment Theory

The Legal Prerogatives of the Tax Administration

Survey of French Finance

Economic Survey of French Overseas Territories

Book Reviews

Bibliography

French and Foreign Financial Information

C. S. Shoup (New York)

P. Maillet (Paris)

J. Sykes (Exeter)

P. Lavigne (Strasbourg)

G. Rullière (Lyon)

G. Tixier (Toulouse)

The Review is published quarterly. Yearly Subscription abroad, 4,000 French francs or 9.00 U.S.A. dollars. Orders can be placed to Pichon and Durand-Auzias, 20 rue Soufflot, Paris (V°), and through Stechert-Hafner, 31 East 10th Street, New York, N. Y.

PUBLIC FINANCE / FINANCES PUBLIQUES

International quarterly journal devoted to the study of fiscal policy and related problems

Publisher and Editor in Chief: DR. J. A. MONOD DE FROIDEVILLE

Emmalaan 2, Haarlem, Netherlands

VOLUME XII

1957

NUMBER 3

Maurice Masoin, Brussels. The Financing of Social Security

Federico Julio Herschel and Samuel Itzcovich,

Buenos Aires. Fiscal Policy in Argentina (Continuation)

D. J. Morgan, Jamaica. Land Valuation and Land Taxation in Jamaica

Schlomo Riemer, Jerusalem. Inflation in Israel

Contributions in French or German are followed by an English Summary.

Subscription Price \$8.75 a year, postage free.

K Y K L O S

INTERNATIONAL REVIEW FOR SOCIAL SCIENCES

VOL. X — 1957 — FASC. 4

Approaches to the Integration of Social Inquiry: A Critical Evaluation

Monetary Policy and the Public Debt: An Appraisal of Post-War Developments

in the U.S.A.

Le facteur humain et les activités modernes des pays sous-développés —

l'exemple de l'A.O.F.

A Nineteenth Century Approach to a West European Common Market

"Individualismus" oder "Menschlichkeit"?

A French Contribution to the Problem of "Underdevelopment"

Masse gegen Demokratie

Profit and Saving

K. William Kapp

John H. Kareken

Marcel Capel

W. O. Henderson

Leopold v. Wiese

N. Koestner

Carl J. Friedrich

Hans Neisser

BOOK REVIEWS — TITLES OF NEW BOOKS

All articles are followed by a summary in English, French and German. Published quarterly. Annual subscription Sw.Fr. 26. — (US \$6.00), postage included. Order from your bookstore or the publisher: Kyklos-Verlag, Postfach 20 365, Basel 2, Switzerland.

Towards a More General Theory of Value

by **EDWARD HASTINGS CHAMBERLIN**, David A. Wells Professor of Political Economy, Harvard University.

This book brings together four new essays and a number of recent articles by Chamberlin, constituting the further development of his theory and its application to specific economic problems. **TOWARDS A MORE GENERAL THEORY OF VALUE** is designed for the general economist, and also to serve as a text or supplement for courses in economic theory, advanced theory, and various applied fields, including marketing and industrial organization. 1957 326 pages \$5.00

Essays in the Theory of Economic Growth

by **EVSEY D. DOMAR**, Professor of Political Economy, The Johns Hopkins University. 1957 282 pages \$4.50

Outline of Monetary Economics

by **A. C. L. DAY**, Reader in Economics, London School of Economics and Political Science. 1957 574 pages \$7.00

OXFORD UNIVERSITY PRESS, Inc., NEW YORK

American Economic Review

CONTENTS

VOLUME XLVII	DECEMBER 1957	NUMBER 6
--------------	---------------	----------

ARTICLES

Monetary Systems and Accelerator Models	<i>H. P. Minsky</i>
Wages and Interest: Marxian Economic Models	<i>P. A. Samuelson</i>
Instalment Credit and Prosperity	<i>Alain Enthoven</i>
New Zealand's Experiment in Economic Planning	<i>J. B. Condliffe</i>
Effects of Consumer Attitudes on Purchases	<i>Eva Mueller</i>

REVIEW ARTICLES

Consumer Instalment Credit	<i>W. L. Smith</i>
Professor Cole's "History of Socialist Thought"	<i>P. M. Sweezy</i>

COMMUNICATIONS

External and Internal Public Debt	<i>J. M. Buchanan</i>
Acceleration and Magnification	<i>M. H. Peston</i>

The AMERICAN ECONOMIC REVIEW, a quarterly, is the official publication of the American Economic Association and is sent to all members. The annual dues are \$6.00. Address editorial communications to Dr. Bernard F. Haley, Editor, American Economic Review, Stanford University, Room 220, Stanford, California; for information concerning other publications and activities of the Association, communicate with the Secretary-Treasurer, Dr. James Washington Bell, American Economic Association, Northwestern University, Evanston, Illinois. Send for information booklet.

Recruiting Telephone Ideas for the Future



What will the telephone of the future be like? Key members of CPPD discuss some possible models. Will they work? Are they marketable? Will they stand up?

Bell System's new Customer Products Planning Division has the fascinating job of generating, screening and testing new ideas for ever-better telephone equipment and service.

Here in this quiet room is shaped an important part of the future of the telephone.

For here are gathered together from many sources the hundreds of new engineering and styling ideas . . . even the "screwball notions" . . . from which the telephone of tomorrow will be developed.

Which are good? Which are bad? It is the responsibility of the Customer Products Planning Division to find out. And to select for development and production those items that people really want.

No idea seems too farfetched for careful consideration by this hard-headed but hopeful group.

They go on the premise that even a poor idea may spark a good one, and that you never know how good an idea is until you try it.

So, when an idea looks promising, working models are developed and designed by the Bell Telephone Laboratories, built by Western Electric Company, and tried out in homes or offices. Thus no bets are missed, and no costly mistakes are made.

This is just one reason for the success of Bell System's program of research for ever-better telephone service.

Bell Telephone System

Working together to bring people together



ECONOMIC DEVELOPMENT

By CHARLES P. KINDLEBERGER, Massachusetts Institute of Technology. *McGraw-Hill Economics Handbook Series*. 325 pages, \$6.50

A senior-graduate level college text dealing mainly with current economic development problems. The author shows how the various development theories which emphasize variables are interrelated. Special stress is given to questions of the relation of market growth and the money mechanism to productivity, and to the international aspects of contrived development.

BUSINESS CYCLES AND ECONOMIC GROWTH

By JAMES S. DUESENBERY, Harvard University. *McGraw-Hill Economics Handbook Series*. Ready in April

This timely new economics text offers a general explanation of the causes of income growth and fluctuation. It includes a thorough explanation of income movement and its components with a simplified description of economic behavior. The theory of income movement provides a basis for analysis of the impact of structural change on the movement of income. Chapters on "Marginal Investment Efficiency"; "Oligopoly and Investment"; and "Wages and Labor Supply" are included.

ECONOMICS: An Introductory Analysis

By PAUL A. SAMUELSON, Massachusetts Institute of Technology. New Fourth Edition. Ready in March

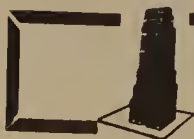
A newly revised edition of this famous beginning economics text. In a completely new design and two-color format, it places greater emphasis on inflation and monetary policy, and on the microeconomics of pricing. The author concentrates on the vital problems of inflation and depression; family and national income; prices, wages, and profits; and conditions leading to rapid economic progress and security. Even more difficult material seems easier in this well-written text.

LABOR

By NEIL W. CHAMBERLAIN, Director, Program in Economic Development and Administration, Ford Foundation. Ready in May

Emphasis throughout this outstanding new text is on the economic effects of collective bargaining. Using both *fact* and *theory*, it builds a conception of the total role of organized labor in the economy. Its main concern is the effects of unions on the economic activity of the U. S. The book has two general parts: an analysis of the people and institutions involved; and, an examination of the impact which organized workers have had on the economy.

SEND FOR COPIES ON APPROVAL



McGRAW-HILL BOOK COMPANY, INC.

330 West 42nd Street

New York, N. Y.

a clear and effective presentation

MONEY AND BANKING

by Leland J. PRITCHARD, University of Kansas

A careful balance of theory and application, this new text introduces students in both business and liberal arts courses to the principal aspects of bank administration and to the "economics" of money and banking. By treating the intermediary financial institutions concisely, the author has been able to cover more fully the commercial and reserve banks and the areas of theory and international finance. Dr. Pritchard's twenty-three years of teaching experience provide the background for a clear and penetrating examination of the nature and origin of money; the relationship of money to prices, production and employment, both in domestic and in the international spheres; and the effects of governmental monetary and financial policies on the economy.

A Spring Publication.

Houghton Mifflin Company

*Boston 7
New York 16
Chicago 16
Dallas 1
Palo Alto*

THE WORLD'S SUGAR Progress and Policy

Vladimir P. Timoshenko and Boris C. Swerling. A wide-ranging study of the evolution of the international sugareconomy which emphasizes the interrelationships between public policy and economic development. *A publication of the Food Research Institute, Stanford University.* \$6.50

THE WORLD FERTILIZER ECONOMY

Mirko Lamer. This first world-wide study of the economics of commercial fertilizers is both a comprehensive analysis and an invaluable reference. *A publication of the Food Research Institute, Stanford University.* \$12.50

THE COMBINED FOOD BOARD A Study in Wartime International Planning

Eric Roll. An account of the origin, growth, influence and demise of the supreme Allied body dealing with international production, distribution, and consumption of foodstuffs during the years 1942-1946. *A publication of the Food Research Institute, Stanford University.* \$7.50

THE ECONOMY OF PAKISTAN

J. Russell Andrus and Azizali F. Mohammed. An authoritative study which considers all aspects of the economy of Pakistan. \$8.50

*Stanford University Press
Stanford, California*

PRINCETON UNIVERSITY PRESS

Capital in Agriculture

Its Formation and Financing since 1870

By Alvin S. Tostlebe

This book measures the growth of physical farm capital for an 80-year span, and the growth of all financial assets used in farming since 1900. It relates this growth to the farm labor force and to output, and indicates the principal determinants of investment and the sources of financing that made the investments possible. *Published for the National Bureau of Economic Research.*

250 PAGES. MAPS AND CHARTS. \$6.00

Organized Business in France

By Henry W. Ehrmann

The first book to give a comprehensive description of a major French pressure group — the employers' trade and association movement. Professor Ehrmann encompasses a dramatic period of French history from 1936 to the present, when organized business played a prominent role, to show how this movement operates, how its policies are formed, and how its influence extends to the heart of French politics. He examines the chances for economic growth in mid-century France and analyzes the economic and social concepts of organized business as expressed in policy statements and in daily practice. 480 PAGES. \$7.50

Financial Intermediaries in the American Economy since 1900

By Raymond W. Goldsmith

A unique treatment of the growth of financial institutions since the turn of the century and of their role in the American economy. The author shows the increase in the practice of entrusting savings to institutions, measures the share of the nation's assets held by financial intermediaries, and discusses the implications of the present trend. *Published for the National Bureau of Economic Research.*

420 PAGES. CHARTS AND TABLES. \$8.50

Order from your bookstore, or

Princeton University Press, Princeton, N. J.

THE INTERNATIONAL ECONOMY, *Revised Edition*

by PAUL T. ELLSWORTH, *University of Wisconsin*. Condensed and brought thoroughly up to date, this complete revision of an outstanding text highlights a more rigorous, systematic presentation of the theory of international economics. *Spring 1958*

MANAGEMENT PRINCIPLES AND PRACTICES

by DALTON E. MCFARLAND, *Michigan State University*. Drawing upon his industrial and academic experience, the author stresses the importance of social science research in all phases of management. Study questions, case materials, and reference guides supplement the text. *Spring 1958*

ESSENTIALS OF ACCOUNTING

by WILLIAM A. PATON *and* ROBERT L. DIXON, *University of Michigan*. Providing a more readable, teachable treatment of accounting fundamentals, this up-to-date volume incorporates new material and new emphasis on the management point of view and on the importance of cost accounting. *May 1958*

INTERNATIONAL ECONOMIC PAPERS No. 7

This is the new seventh volume in an annual series containing translations of articles of world-wide economic interest. An important addition to the series, this collection features contributors well-recognized in their fields. *Spring 1958*

PROBABILITY, STATISTICS AND TRUTH

by RICHARD VON MISES. *Revised by* HILDA VON MISES. The new English translation of one of the internationally outstanding works on statistics, this is an important analysis of the quantitative concept of probability in probability theory. *1957, 244 pages, \$5.00*

The Macmillan Company

60 FIFTH AVENUE, NEW YORK 11

THE QUARTERLY JOURNAL OF ECONOMICS

FOUNDED 1886

Vol. LXXIII

May, 1958

No. 2

A LETTER FROM ADAM SMITH.....	Z. CLARK DICKINSON	157
IS THE INCOME OF THE "CADRES" A SPECIAL CLASS OF WAGES?....	JEAN MARCHAL and JACQUES LECAILLON	166
INTERNATIONAL SPECIALIZATION AND THE CONCEPT OF BALANCED GROWTH.....	JOHN SHEAHAN	183
WHAT IS THE ECONOMIC SYSTEM?.....	DAVID McCORD WRIGHT	198
THE CLASSICAL ECONOMISTS AND THE FACTORY ACTS — A RE-EXAMINATION.....	MARK BLAUG	211
ADAPTIVE EXPECTATIONS AND COBWEB PHENOMENA MARC NERLOVE		227
COMPETITION, CREDIT POLICIES, AND THE CAPTIVE FINANCE COMPANY.....	PAUL H. BANNER	241
LOCATION AND THE THEORY OF PRODUCTION..	LEON N. MOSES	259
THE EFFECT OF DEVALUATION ON THE PRICE LEVEL EGON SOHMEN		273
REVIEWS, NOTES AND DISCUSSIONS		
The Business Cycle in the Post-War World: A Review.....	James Tobin	284
Investments in United States Government Securities by Nonfinancial Corporations, 1952-1956:		
Comment.....	Colin D. Campbell	292
Reply.....	Hyman P. Minsky	297
Tariffs and the Balance of Payments:		
Comment.....	Harry G. Johnson	301
Reply.....	S. A. Ozga	302
Recent Publications.....		304

PUBLISHED BY HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS

The Quarterly Journal of Economics

Published by Harvard University

Editor

EDWARD H. CHAMBERLIN

Associate Editors

ABRAM BERGSON
JOHN D. BLACK*
J. K. BUTTERS
T. N. CARVER*
ROBERT DORFMAN
JAMES S. DUESENBERY
JOHN T. DUNLOP
JOHN KENNETH GALBRAITH
ALEXANDER GERSCHENKRON
GOTTFRIED HABERLER
ALVIN H. HANSEN*

SEYMOUR E. HARRIS
CARL KAYSEN
WASSILY LEONTIEF
EDWARD S. MASON
GUY H. ORCUTT
SUMNER H. SLICHTER
ARTHUR SMITHIES
OVERTON H. TAYLOR
ABBOT P. USHER*
JOHN H. WILLIAMS*
EDWIN B. WILSON*

* Emeritus

Assistant Editor

CATHERINE RUGGLES GERRISH

Books, periodicals, and manuscripts to be addressed, THE QUARTERLY JOURNAL OF ECONOMICS, M-12, Littauer Center, Cambridge 38, Mass.

Business letters to be addressed, HARVARD UNIVERSITY PRESS, 79 Garden Street, Cambridge 38, Mass.

Subscriptions \$5.00 a year. Three years, \$14.00. Single copies (except those published in two parts) \$1.50, postpaid. A special subscription rate of \$3.00 is offered to students. Such special subscriptions may be for one, two, three, or four years. Those desiring to take advantage of this reduced rate should include with their remittance a statement from an officer of their college or university that they are *at that time* "regularly enrolled students who have not yet received the Doctor's degree."

CONTENTS FOR FEBRUARY, 1958

THE STRUGGLE FOR COMMERCIAL BANK SAVINGS

David A. Alhadeff and Charlotte P. Alhadeff

FORECASTING USES OF ANTICIPATORY DATA ON INVESTMENT AND SALES

F. Modigliani and H. M. Weingartner

ECONOMIC PLANNING RECONSIDERED..... John E. Elliott

ON THE THEORY OF NEGOTIATION..... Carl M. Stevens

MONETARY SURVEYS AND MONETARY ANALYSIS..... Claudio Segré

THE FUNCTION OF WAGE POLICY: THE AUSTRALIAN EXPERIENCE..... J. E. Isaac

MOTOR CARRIER COSTS AND MINIMUM RATE REGULATION..... Howard W. Nicholson

Recent Publications

COPYRIGHT, 1958, BY THE PRESIDENT AND FELLOWS OF HARVARD COLLEGE

Entered as Second-Class Mail Matter. Acceptance for mailing at special postage rate provided for by Section 1103, Act of October 3, 1917, authorized on July 31, 1918

Published by the Harvard University Press, Cambridge, Mass., U.S.A.

London: Oxford University Press

THE QUARTERLY JOURNAL OF ECONOMICS

Vol. LXXIII

May, 1958

No. 2

A LETTER FROM ADAM SMITH

By Z. CLARK DICKINSON

Correspondence and other manuscripts tending to illuminate Smith's life, thought, and activities are widely scattered; each fragment a collector's item. As is well known, the great philosopher-economist caused to be destroyed many of his unpublished scripts; apparently including some letters he had received and preserved. As to his outgoing mail, however, there was probably no serious loss due to his mandate, since he almost never retained copies of letters sent. And he was not addicted to letter-writing; so that, compared with correspondence of some other notable persons, Smith's letters were rather few and short.

His principal biographer remarks, apropos a note from Smith to Lord Shelburne, that it was "written in the first instance, like many others of Smith's extant letters, to do a service to a friend."¹ In this instance and in some others — notably as to Smith's relations with Charles Townshend and the latter's stepson, the young Duke of Buccleuch — Rae was able to find theretofore unpublished correspondence; but to a greater extent he used published biographical materials, such as letters of David Hume.

Only a year later than the appearance of Rae's biography saw Cannan's publication, with editorial matter, of a newly-discovered script, since commonly referred to as Smith's *Glasgow Lectures* and still believed to be a student's notes indicating an early stage of the development of Smith's economic views.²

1. John Rae, *Life of Adam Smith* (London, 1895), p. 235. The letter there referred to was dated 12 February, 1767; Rae found it among Lansdowne Mss., of which more will be said below.

2. The title page and another specimen page of this manuscript, now in the Glasgow University Library, are photographically reproduced on pp. 388, 389 of William R. Scott, *Adam Smith as Student and Professor* (Glasgow, 1937). The handwriting appears to be that of a professional scribe, as is not inconsistent with the date "MDCCLXVI" on the title page — some two years after Smith resigned from Glasgow.

Through following decades other Smithiana have been turned up, especially by the sagacity, energy, and devotion of William Robert Scott (1868–1940) who, when he published the magnificent work mentioned above and hereinafter, was Adam Smith Professor of Political Economy at Glasgow. A major find, in the Townshend papers of the Buccleuch family, Scott designated “An Early Draft of Part of the *Wealth of Nations* (c. 1763)”; it shows numerous parallels with the great book which are not in the *Glasgow Lectures*. Scott found some other materials pertaining to Smith among the Townshend-Buccleuch papers, mostly at the Dalkeith House seat; one group was used in Scott’s note “Adam Smith at Downing Street, 1766–67.”³

Scott’s principal resources of unpublished manuscripts were at his own Glasgow University, many of them in the Bannerman papers, coming from Smith’s kinsmen descended from his nephew-heir, David Douglas, later Lord Reston. And Scott’s other acknowledgments make almost a “Who’s Who” of scholarly resources on Smith, including manuscripts at Moscow and Tokio and a group of Smith letters belonging to Professor Jacob Hollander.

Some years ago I learned that the William L. Clements Library of the University of Michigan had come into possession of a Smith letter, found in the mass of Shelburne papers purchased by Mr. Clements about 1920 in the course of his shopping for rare Americana. Only very recently, through a note in the “Downing Street” article by Scott, did I become aware that the Clements Library also holds a memorandum, almost certainly composed by Adam Smith for the North administration, endorsed as “Smith’s Thoughts on the Contest with America; Feby. 1778” by Smith’s friend Alexander Wedderburn (1733–1805), Solicitor-General or Attorney-General.⁴

The letter of Smith referred to in the preceding paragraph, almost certainly directed to Lord Shelburne and written a little less

3. *Economic History Review*, Vol. 6 (Oct. 1935), pp. 79–89. This episode in Smith’s life shows him serving Townshend (Chancellor of the Exchequer) as brain-truster on a projected history of the British sinking fund (for reducing national debt). Presumably following Townshend’s death in 1767, Smith seems to have removed for his own use the materials he had worked up; but a few were overlooked and remained in the Townshend papers.

4. See note by G. H. Guttridge, in *American Historical Review*, Vol. 38 (July 1933), pp. 714–20, which prints the whole memorandum and cites parallel “thoughts” of Smith in his lengthy section on colonies, in the *Wealth of Nations*. The memorandum, in the hand of a professional copyist, is in the Clements Library’s Wedderburn papers. (Wedderburn became Baron Loughborough in 1780; Earl of Rosslyn, 1801.) Incidentally, the Clements Library also has a French reprint of Smith’s section on colonies — *Fragments sur les Colonies en Général, et sur celles des Anglois en Particulier* (Lausanne: chez La Société Typographique, 1778).

than a year later than his note to the same personage, cited herein from Rae's *Life*, is now published through courtesy of the William L. Clements Library.

A few comments will clarify some parts of the letter. First, during the year 1766–67, as told above, Smith collaborated with Townshend on the history of the sinking fund, probably beginning soon after his return with the Duke from France in November 1766. During the same winter he also did some research for Shelburne (with whom his contacts began in 1759) on the history of the Roman colonies.⁵

Especially between 1763 and Townshend's death in September 1767, this statesman and Shelburne were prominent political antagonists with reference to the American colonies; Townshend earning enmity in the colonies by supporting the stamp tax and other financial measures; Shelburne, first as president of the Board of Trade and in 1766–67 as Secretary of State with jurisdiction over these colonies, proponent of a lighter rein and, eventually, of something not altogether unlike the later dominion-status idea.

From Kirkaldy on 7 June 1767 Smith wrote to Hume "to recommend . . . the Count de Sarsfield, the best and most agreeable friend I had in France . . ." ⁶ Evidently about the same time Smith commended the Count to Shelburne, either orally or by a note which has not come to light.⁷

The topic other than Sarsfield which catches our eye in the letter to Shelburne dated 27 January 1768 is "a change . . . in the original contract I made with the Duke of Buccleugh [usual spelling at that time] . . ." Again Rae is of some help, notably as to the Duke's marriage, autumn of 1767, after Townshend's death, and the avuncular position occupied by Adam Smith for weeks on end at the ducal seat, Dalkeith House.⁸ Smith and the new Duchess were fortunately thoroughly congenial; in fact, ten years later she joined her husband

5. See his letter to Shelburne of 12 February 1767, printed in Rae's *Life*, pp. 235, 236; also adjacent pages for support and elaboration of some points in my text.

6. This letter printed in Rae, *Life*, pp. 241, 242. Adjoining pages give more on Sarsfield, including reference to John Adams' favorable impression of him; his association with Turgot, etc.

7. My colleague Professor Verner W. Crane comments here: "The Comte de Sarsfield reminded Franklin, soon after the latter arrived in France, that they had met in England; and thereafter saw a good deal of him during his diplomatic mission. There are numerous Sarsfield letters in the Franklin papers, American Philosophical Society." As indicated further below, Shelburne had a number of associations with Franklin and other philosophers, British and French, of that period.

8. *Life*, pp. 242 ff.

Kirkaldy 27 Jan: 1768:

My Lord

I should have written to your Lordship a long time ago, to thank you for the very great kindness which you was so good as to shew me when I was last at London; but I flattered myself that your Lordship knew me too well to need to be assured of my sense of it by letter; & the little transactions of this Country afford nothing that ^{can} furnish any amusement to your Lordship. I now write to your Lordship to thank you for your kindness to another man, I mean my very worthy & excellent friend the Count de Sarsfield. He writes me that you have been excessively good to him; & I have the vanity, which perhaps you will laugh at, to impute some part of the kindness which you have shewn to him, to the regard with which you have been so good as to honour me. There is nothing that your Lordship could Possibly have done that would have bound me more effectually to you. I have no doubt but you

Kirkaldy 27 Jan: 1768:

MY LORD

I should have written to your Lordship a long time ago, to thank you for the very great kindness which you was so good as to shew me when I was last at London; but I flattered myself that your Lordship knew me too well to need to be assured of my sense of it by letter; & the little transactions of this Country afford nothing that can furnish any amusement to your Lordship. I now write to your Lordship to thank you for your kindness to another man, I mean my very worthy & excellent friend the Count de Sarsfield. He writes me that you have been excessively good to him; & I have the vanity, which perhaps you will laugh at, to impute some part of the kindness which you have shewn to him, to the regard with which you have been so good as to honour me. There is nothing that your Lordship could Possibly have done that would have bound me more effectually to you. I have no doubt but you have found him the same plain, worthy, sensible man which I described him to be.

have found him the same plain, worthy, sensible man which y^e described him to be.

Since y^e came to this country y^e have employed my-
self pretty ^{much} in the manner that y^e proposed. I have not
however, made all the Progress that y^e expected;
I have resolved, therefore, to prolong my stay here till
November next, perhaps, till after the Christmas
holidays next winter.

When your Lordship sees Coll: Clerk I beg that
you would tell him that y^e ^{have} followed his advice
exactly with regard to a change which he proposed.
I should make in the original contract y^e made
with the Duke of Buccleugh. I am very much
obliged to him for his counsell & I feel the good
effects of it every day. He will explain this to
your Lordship by word of mouth much better than
I could by a long letter written in so very bad a
hand.

I beg to be remembered in the most respectful
manner to Lady Shelburne & that your Lord-
ship would believe me to be with the highest respect
& esteem My Lord Your Lordships most obliged
& devoted Servant
Adam Smith

THIS PHOTOSTAT IS FOR REFERENCE ONLY. PERMISSION TO
REPRODUCE, IN FULL OR IN PART, MUST BE SECURED FROM
THE WILLIAM L. CLEMENTS LIBRARY, ANN ARBOR, MICH.

Since I came to this country I have employed myself pretty much in the
manner that I proposed. I have not, however, made all the Progress that I
expected; I have resolved, therfor, to prolong my stay here till November next,
perhaps, till after the Christmas holidays next winter.

When your Lordship sees Coll: Clerk I beg that you would tell him that
I have followed his advice exactly with regard to a change which he proposed I
should make in the original contract I made with the Duke of Buccleugh. I am
very much obliged to him for his counsell & I feel the good effects of it every day.
He will explain this to your Lordship by word of mouth much better than I could
by a long letter written in so very bad a hand.

I beg to be remembered in the most respectful manner to Lady Shelburne
& that your Lordship would believe me to be with the highest respect & esteem
My Lord Your Lordships most obliged

& devoted Servant

ADAM SMITH

in urging upon the North government appointment of Smith as Commissioner of Customs at Edinburgh.⁹ The revision of terms mentioned might be implied by Scott's finding that "The terms of his engagement [to be the young Duke's companion] were more favorable to him [Smith] than those which have been recorded. It appears from one of his receipts at Dalkeith House that he received £500 a year while he was travelling with the Duke and an annuity of £300 a year."¹ Rae had said "The terms were a salary of £300 a year, with travelling expenses while abroad, and a pension of £300 a year for life afterwards."² Conceivably, the change referred to by Smith was a commutation of the salary-and-expense-allowance to £500 a year for all the time they were abroad.

About 1780, in a letter relative to the Danish translation of his *Wealth*, Smith gave one of the various testimonials which he and the Duke wrote as to their continuing mutual high regard. ". . . By the interest of the Duke of Buccleugh," he wrote this Danish correspondent, "I was appointed to my present [Commissioner of Customs] office; which, though it requires a good deal of attendance, is both easy and honourable, and for my way of living sufficiently beneficial. Upon my appointment I proposed to surrender the annuity which had been settled upon me by the Tutors of the Duke of Buccleugh . . . and which had been renewed by his Grace when he became of age, . . . But his Grace sent me word . . . that though I had considered what was fit for my own honour, I had not considered what was fit for his, and that he would never suffer it to be suspected that he had procured an office for his friend, in order to relieve himself . . . of such an annuity . . ." (It seems likely that Wedderburn, and perhaps other admirers and friends of Smith, had at least as much influence toward his Commissioner appointment as had the Duke.)

The Buccleuchs evidently expressed their appreciation and friendship mainly in other ways than letter-writing; and it is pleasant to read the little note from the Duke, written but a few months before Smith's death, concluding: "We have long lived in friendship, uninterrupted for one single moment, since we were first acquainted."³

From these Scottish scenes let us turn to the Shelburne family, in relation to Adam Smith. The "Lord Shelburne" referred to above was William Fitzmaurice (1737-1805) who on the death of his father John (in 1761) became Earl of Shelburne; and in 1785 first Marquess

9. Scott, *Adam Smith as Student and Professor*, p. 274.

1. *Ibid.*, p. 97.

2. *Life*, p. 165.

3. These two letters in Scott, *Adam Smith as Student and Professor*, pp. 281, 282, 311.

(or Marquis) of Lansdowne, retaining as a major seat his estate of Bowood in Wiltshire, where he amassed an extensive library of historical manuscripts in addition to his own numerous papers produced by a lengthy and prominent political career. He was a great-grandson of Sir William Petty, succeeding to what was left of the Petty estates in Ireland and England.⁴

In November 1758, about a year before publication of the *Moral Sentiments*, Smith learned from a fellow-Scot (Gilbert Elliot) of Smith's age but apparently not well acquainted with him, that Elliot had recommended to our Lord Shelburne-to-be (then William, Lord Fitzmaurice) that William's younger brother Thomas should be put in Smith's care as a student at Glasgow. William was then aged twenty-one and just out of Oxford; after two years with Smith, Thomas also went to Oxford and also studied under Blackstone. With Elliot's notification there began the friendly association of Smith with the Shelburnes of which the tone of the letter of 27 January 1768, printed above, is symptomatic. Scott has found and printed much of the correspondence involved, mostly reports of Smith to John, Lord Shelburne, concerning his younger son Thomas.⁵

"Our" William Fitzmaurice, Earl of Shelburne, rapidly became a power in British politics. When aged twenty-six he became president of the Board of Trade (1763), and thus began his immersion in British relations with the American colonies which proceeded through service as a Secretary of State and later prime minister (1782). During brief periods of relative inactivity in politics he cultivated friendship with a number of eighteenth century philosophers — including Franklin, Bentham, Hume, Priestley (Shelburne's librarian at Bowood for some years). His most intimate French friend over many years was the Abbé Morellet, whose *Letters to Lord Shelburne* have been edited and published by Edmond, Lord Fitzmaurice.

I have told above that some written communications of Smith to Shelburne are extant; and doubtless more will be brought to light, especially as the very numerous Shelburne-Lansdowne papers are worked over further. Another allusion to the association was made by Smith in a note written to the Abbé Morellet in May 1786. "After

4. A *Life* of him ("with extracts from his papers and correspondence") was published by his descendant, Lord Edmond Fitzmaurice, in 1875 and 1876; republished in two volumes with some revisions 1912.

5. *Adam Smith as Student and Professor*, pp. 239-54. Scott points out that Smith was also meticulous in reporting frequently to Townshend concerning the Duke; and some passages of his stepfather's letters to the Duke concerning Smith and advice to the youth are given by Scott in the "Smith at Downing Street" article — see above.

so long an Interruption of our correspondence," this letter began, "I should have been afraid to put you in mind of an old acquaintance, if I had not understood from our most valuable friend the Marquis of Lansdown, that you still did me the honour to remember me with some degree of kindness."⁶

But why are written communications from Shelburne to Smith so scanty — if any? An easy hypothesis is that they were destroyed along with other Smith papers; yet somehow a rather lengthy and very interesting letter apropos Thomas, from his father John, Lord Shelburne, survived. "Our" Lord Shelburne (William) was ready enough to write lengthy political letters, well phrased and meaty; the stout volumes of the Fitzmaurice biography are half-filled with some of them. This biography, preoccupied as it is with Shelburne-Lansdowne's parts on the political stage, has remarkably little to say of Adam Smith; yet what it does say tends to bear out in spirit the tradition, passed along by Rae from Dugald Stewart, that Shelburne said, "I owe to a journey I made with Mr. Smith [apparently in 1761, in connection with putting Thomas Fitzmaurice in Smith's care] . . . the difference between light and darkness through the best part of my life. The novelty of his principles, added to my youth and prejudices, made me unable to comprehend them at the time, but he urged them with so much benevolence, as well as eloquence, that they took a certain hold which . . . I can truly say has constituted ever since the happiness of my life, as well as the source of any little consideration I may have enjoyed in it."

A few adverse comments by Smith on political activities of Shelburne are recorded or implied. Apparently apropos the latter's rather youthful entente with Lord Bute, Smith had written something to Hume so that the latter responded "I see you are much incensed

6. *Ibid.*, pp. 298, 299. This note illustrates further Rae's generalization that when Smith wrote to anyone, it was likely to be in behalf of some friend — and so does the reply to him from the Duke in 1790 — excerpt quoted above. Smith also condoles the Abbé "on the many heavy losses which the Society, in which I had the pleasure of seeing you about twenty years ago, have sustained through the death of so many of its greatest ornaments, of Helvetius, of Mr. Turgot, of Mademoiselle D'Espinasse, of Mr. D'Alembert, of Mr. Diderot . . ."

The letter, and Scott's footnotes, contain other interesting points, e.g., that Smith's friend here commended to Morelet was Edinburgh professor John Bruce, who was taking abroad Robert Dundas, "son," as Smith put it, "of the gentleman who may be regarded as our present Minister for Scotland." This father was Henry Dundas, who became Viscount Melville. Scott remarks, in line with a previous parallel observation, "Adam Smith was fortunate in his relations with the two most influential men in Scotland during the eighteenth century. John, Duke of Argyle, was his early patron and he enjoyed the friendship of Henry Dundas."

with that nobleman [Shelburne] but he always speaks of you with regard.'"⁷ And in 1782 Smith expressed opposition to Shelburne's assuming the premier cabinet post instead of going along with other Whigs who hoped further to clip the King's wings.⁸ In such reactions Smith seems somewhat naive and unappreciative of the extent to which Shelburne was attempting to give practical effect to principles taught by Smith.

Among autobiographical fragments which he penned in his latest years, Shelburne wrote (1801) "Professor Adam Smith's principles have remained unanswered for above thirty years, and yet when it is attempted to act upon any of them, what a clamour!"⁹ Another bit of writing by Shelburne in the same period (1802) to his long-time friend, the Abbé Morellet, remarked: "I have not changed an atom of the principles I first imbibed from you and Adam Smith. They make a woeful slow progress, but I cannot look upon them as extinct; on the contrary they must prevail in the end like the sea . . ."¹¹

It seems entirely probable that Smith was warmly welcome in Shelburne's household on the few occasions when it was convenient for both to be together. Smith maintained some contact with a London literary club, which included Dr. Johnson, but on various accounts he tended to stay in Scotland, where probably he visited from time to time in the Buccleuch household at Dalkeith House.

However well these surmises may be borne out by other discoveries and researches, the letter here printed tends to confirm other evidences that Smith and Shelburne, whose views and influences were so nearly alike on so many matters, had personal contacts which were warm though not numerous.

Z. CLARK DICKINSON.

UNIVERSITY OF MICHIGAN

7. Rae, *Life*, p. 162. This incident was in 1763.

8. *Ibid.*, pp. 378, 379.

9. Lord Fitzmaurice, *Life of William, Earl of Shelburne*, Vol. 1, pp. 18, 19. (2d and rev. ed., 1912.)

1. *Ibid.*, Vol. 2, pp. 430, 431. Biographer Fitzmaurice remarks and quotes (p. 308) on Shelburne's "hatred of the Scotch," expressed, as Fitzmaurice says, "in words almost worthy of Dr. Johnson when dictating on the same subject . . ." Shelburne, however, like Johnson took kindly to the Scot Boswell; also to Hume; and Richard Oswald, a prominent Scottish businessman who had been introduced by Smith to Shelburne, became the latter's most trusted negotiator with American commissioners in 1782. (*Ibid.*, Vol. 2, p. 119.) It is most unlikely that the anti-Scot prejudice noted by his biographer cooled Shelburne's warm regard for Smith.

IS THE INCOME OF THE "CADRES" A SPECIAL CLASS OF WAGES?

(A STUDY IN THE LIGHT OF FRENCH EXPERIENCE)

By JEAN MARCHAL AND JACQUES LÉCAILLON

TRANSLATED BY M. BRONFENBRENNER

Introduction, 166. — I. Distinctive characteristics of the "cadres," 167. — II. Income of the "cadres" relative to other wages, 171. — III. Causes for the autonomy of the "cadres" in the modern French economy, 178.

At the present time, French economists use the word "cadres" to designate a category of salaried employees. This usage is recent. In the 1936 edition of the *Petit Dictionnaire Larousse*, one finds the following definitions under the word "cadre":

"A frame of wood or bronze surrounding a mirror or picture; a framework supporting the casing of a well; a frame placed in beehives; a frame used as a suspended berth. Figuratively: the limits enclosing a space; the space itself; a plan of a work of the mind; a group of petty or noncommissioned officers forming the nucleus of a military body."

It is clearly from this last meaning that the new one has arisen. Let us therefore add at once:

"A group corresponding to the noncommissioned officers of a body of workers, forming its nucleus — more precisely, a group of salaried 'noncommissioned officers.' "

So far as we know, there exists no corresponding term in English or German. There is, of course, in English the word "management," but this word covers only the higher "cadres" to the exclusion of the middle grades. Furthermore, it includes entrepreneurs who are not salaried employees at all. As for the English word "professional," it denotes specialists who are as likely to be working for themselves as for others.¹

This is not to say that the workers designated in France by the term "cadres" have no equivalents in the American, British, or German economies. Such workers exist without doubt, but there is no word to designate them and them alone. Neither within or without the group we call "cadres" has anyone paid sufficient attention to its

1. Translator's note: Other possible English equivalents are the British "salaried staff" and the American "middle management."

role and peculiarities to give it a special name. For several years, however, a considerable number of studies have been published in the United States as well as in France, stressing the importance of "cadres" in production. It has been shown that a shortage of "cadres" may constitute a bottleneck and become a cause of unemployment for ordinary workers.

But nobody, we believe, has stressed sufficiently the peculiarities presented by the income of the "cadres." It is conventional in every country to construct the theory of the distribution of the national income along the traditional lines of wages, interest, rent, and profit, *without setting up subdivisions within these categories*. In a volume currently in press² we propose to discuss this question and to show that the classical categories are by no means homogeneous. It is in our opinion a serious error, *considering the structure of a modern economy*, to believe that the profit of agricultural entrepreneurs is of the same nature (i.e., follows the same principles) as the profit of industrial or commercial enterprises, or that within the industrial and commercial sectors the profit of large corporations behaves in the same way as the profit of individual proprietorships, that the interest on capital lent in the form of money is identical with the return on real capital goods, or finally, that all the remunerations classified as wages are similar to each other.

We do not wish to treat this entire question here, but simply to consider one important point, whether the income of the "cadres" is a wage like other wages, or whether it does not present certain peculiarities and should not constitute one of the subcategories which any theory concerned with realism should substitute for the single category of wages.

I

What are the distinctive characteristics of the "cadres"? To answer, we may consider the problem successively from two points of view, one qualitative and the other quantitative. That is to say, after defining the "cadres" we shall attempt an estimate of their numbers in the contemporary French economy.

A. By what traits may a member of the "cadres" be recognized? Considering the question from strict economic analysis, we think we can identify him by the conjunction of two elements. He is a seller of labor, i.e., a workingman. At the same time, the labor which he

2. Jean Marchal and Jacques Lécaillon, *La répartition du revenu national*, Première Partie: *Les participants*, tome I: "Les salariés," tome 2: "Les non-salariés" (Paris: Librairie de Médecis. 1957-58); Seconde Partie: *Les Processus* (in preparation).

offers for sale has a heightened efficiency, since it has been improved by a process analogous to capitalization.

a. He is first and foremost a seller of labor. He is an individual who rents his labor to the head of an enterprise for a period determined by means of a contractual remuneration. Of course it may happen, and indeed happens more frequently the higher one advances in the hierarchy, that there is combined with this contractual remuneration some participation in the profits of the enterprise. But this participation remains secondary. *The "cadres" are working people.* In this way they may be distinguished not only from proprietors but also from attorneys, physicians, architects, or in general from all who practice liberal professions, enter into contact with clients directly, and operate, as we say, "on their own account."

b. The labor offered on the market by a member of the "cadres" has a heightened efficiency. The reason for this is that he has received a general and technical training superior to those received by most of his fellow-citizens. As a result he is able to fulfill tasks which ordinary folk cannot. These tasks often involve a share in the management, or at least put him into contact with the management.

What is important, we must emphasize, is not the degree of training in itself but the fact that this training is *superior* to that of most citizens, that it gives to the "cadres" particular characteristics, that it makes of the "cadres" a *minority group* within the working class.

We may utilize here the concept of roundabout production introduced into economic science by Böhm-Bawerk. Capitalistic production, in his view, involves a detour in not producing directly the commodities we desire but initially the tools and intermediate products by means of which we ultimately obtain in larger quantity and at lower cost the consumption goods we desire. Capitalistic production requires time.

The member of the "cadres," similarly, is an individual who instead of entering the labor market at the first occasion has had the opportunity, by reason of circumstances to which we shall return later, to spend more time and to take a detour. He has acquired training superior to that of others, which places him in a minority group. When he finally presents himself on the labor market, the labor which he has to offer is simultaneously of *better quality* and of *greater scarcity*. It is a type of labor which in the same way as an object made by machinery, contains not only labor pain but time (training time) greater than that contained in the labor offered by the bulk of his fellow-citizens.

Just as it is difficult to find in modern civilization examples of noncapitalistic production (carried on without tools), so that we must simply contrast highly with slightly capitalistic production, we must in dealing with labor be content with a distinction between that which contains much time and that which contains little. At the present day, with the progress of training, labor has come to contain increasingly more time. The labor of a member of the "cadres" may be characterized as containing more time than the labor of the ordinary worker in the country and at the time under consideration. In other words, the "cadres" are distinguished from other wage earners not by any absolute level of training, but by surpassing the norm or average.

Thus an investigation carried out by the Institut National d'Etudes Démographiques³ shows that in France (1950), of the wage-earning group 68 per cent of the administrative "cadres" and 36 per cent of the industrial and commercial "cadres" had received some higher training, as against only 6 per cent of subordinate functionaries, 2 per cent of clerks, and no manual workers.

If we compare these results with those for nonwage earners, we can see that the "cadres" approximate practitioners of the liberal professions and also businessmen of whom 63 per cent and 24 per cent respectively have received higher training. They are in contrast with retailers and artisans and with working farmers, for whom the percentages are respectively 3 per cent and 2 per cent.

The member of the "cadres" appears therefore as a wage earner who can be distinguished from others by training superior to the average.

Some may perhaps object that we underestimate the importance of ability. The "cadres," they may say, are not made up of men with superior training, but of those with superior natural aptitudes. In support of their position they will point out that some individuals, thanks to their gifts or to favorable circumstances, attain the status of membership in the "cadres" with no more than average training, and that conversely, others who have received such training may decline to the rank of ordinary wage earner.

But the economist who seeks to formulate a theory must deal with the average or typical case and not with the exceptions. In general, it seems clear that the "cadres" do have superior aptitudes. But this is not sufficient in itself for an individual to gain such status, at least in most cases. He must in addition have a sufficiently long and adequate training. It is this which is important.

3. Marcel Brésard, "Mobilité sociale et dimension de la famille," *Population*, July-Sept., 1950, p. 553.

The definition at which we have arrived leads to the recognition of two sorts of "cadres": economic "cadres" in the strict sense, who exercise their functions within business enterprises, and administrative "cadres" whose function is the direction of state enterprises. It therefore differs from the definition of Jacquin⁴ who limits the term to the economic "cadres" exclusively. It has, however, the advantage of covering all the groups enumerated by the Institut National de la Statistique et des Etudes Economiques in its "Code des Catégories professionnelles" (1951) and covering no others.⁵

According to the Institut National de la Statistique et des Etudes Economiques, the following should be included in the "cadres": engineers and technical directors; salaried persons practicing liberal professions — doctors; architects; public accountants; professors; magistrates; administrative and commercial directors; personnel managers.

On the other hand, all nonsalaried persons should be excluded; and among the salaried employees: those in subordinate intellectual positions (e.g., pharmaceutical assistants) or in subordinate technical positions (e.g., draftsmen, bank clerks, shop salesmen, foremen, supervisors, and all manual workers).

It would thus appear that even though the "cadres" exclude manual workers, they should not be confused with clerical workers in any strict sense, nor with technicians, nor with supervisors. Clerks in effect merely carry out orders. Technicians have a certain competence, but it is minor, being limited to material or mechanical operations, and it does not imply authority over others. Foremen and supervisors have such authority, but it extends only to clerks and manual workers, and rests on knowledge acquired by experience. Their administrative functions result in no way from training superior to the average.

B. From the quantitative viewpoint, how large are the "cadres" of the current French economy?

An employment study carried out in February 1956 by the Institut National de la Statistique et des Etudes Economiques permits us to answer this question. As of that date and according to the definitions we have developed, there were approximately 1,200,000 members of "cadres" of middle rank and 600,000 members of "cadres" of superior rank, or a total of 1,800,000. Within the membership of "cadres" of superior rank, however, were included practitioners of the liberal professions.

4. François Jacquin, *Les cadres de l'industrie et du commerce en France* (Paris: Librairie Armand Colin, 1955).

5. Marcel Brichler, "Classification de la population," *Journal de la Société de Statistique de Paris*, April-June, 1956, p. 130.

In percentage terms there is wide variance as between sectors of the economy. In agriculture the percentage is very small, amounting to only 0.2 per cent. In trade, where there are many small proprietorships, it is still only 6.0 per cent. In manufacturing industry it rises to 8.6 per cent, and in transportation to 13.7 per cent, largely due to the railroads (Société Nationale des Chemins de Fer Français). In the administrative and liberal professions, finally, it reaches 39.2 per cent, but we should remember that this figure includes independent practitioners of the professions.

Relatively to the nineteenth century, the size of the "cadres" has increased substantially. Considering only the "cadres" of commerce and industry, and using hypotheses whose approximate character he stresses, Bleton has tried to evaluate the importance of this group at the middle of the nineteenth century.⁶ He estimates that their membership could not have exceeded 120,000, as against 780,000 at the present time. The increase is therefore one of 600 per cent, whereas that of wage earners in manufacturing does not exceed 60 per cent. Modern production is carried on with an increasing proportion of "cadres."

II

Does the income of the "cadres" present any peculiarities relative to other wages, i.e., to other incomes paid in exchange for the sale of labor.

Before proceeding to a quantitative study, it may be desirable to present certain qualitative remarks:

1. The remuneration of the "cadres" is *monthly*, while the wage of most workers is paid at the end of each week and is a function of the number of hours of work furnished. This difference is not without interest. Not only does the member of the "cadres" benefit from a longer notice in case of dismissal, but he has paid vacations whereas other workers regretfully observe that if they take such vacations, there is a reduction in the number of hours which they work and for which they are paid. We should note, however, that clerical employees are also paid by the month.

2. In the second place, there is often added to the contractual remuneration, in constituting the income of the "cadres," some participation in profits and also some payments in kind, such as lodging, use of an automobile, etc. These elements are more important the higher one rises in the hierarchy. Nonexistent in the public sector,

6. P. Bleton, *Les hommes des temps qui viennent* (Paris: Editions ouvrières, 1956).

they have developed above all in the private sector and in the nationalized sector. They tend to tie the "cadres" to their enterprises and to approximate their incomes in some measure to profits. However, should these elements increase to the point of becoming more important than the contractual remuneration, we should no longer be dealing with "cadres" but with "executives."

3. Finally, the income of the "cadres," quite different from the pay of workers or clerks, is considered by its recipients a private matter and its disclosure is most distasteful to them. This is doubtless connected with the fact that this income varies as between individuals; we shall return to this point later on. But the fact itself is certain and highly significant. It demarcates the group, and tends in some sense to make it comparable to nonwage earners and to oppose it to other wage earners.

Having dealt with these preliminary considerations, we may consider two further questions.

A. The first of these is: at what level (relative to other wages) is the income of the "cadres" at the present time?

To answer this question, we have at our disposal statistical information of high quality from examination of the statements called "Etats 1024," by which the heads of enterprises make known to the fiscal authorities the amounts of wages and salaries paid by them. These statistics have only one drawback. While covering the total of wage and salary payments in the private and nationalized sectors, they neglect the officials in the public service.

Given these limitations, it would appear that in 1952 the mean net annual incomes for full-time workers were approximately as follows:

Manual workers	305,000 francs
Clerical workers	374,000
"Cadres" — middle rank	656,000
"Cadres" — superior rank	1,200,000

One is struck by the size of the disparity. On the base of 100 for the manual workers, we may calculate indexes of disparity in the following fashion:

Manual workers	100
Clerical workers	122
"Cadres" — middle rank	215
"Cadres" — superior rank	393

Almost the entire group of manual workers (97.0 per cent) and the great majority of the clerical workers (85.7 per cent) receive less

than 600,000 francs per year, that is to say, less than 50,000 francs per month, while only half the "cadres" of middle rank (46.0 per cent) and a fifth of the "cadres" of superior rank (20.0 per cent) remain below this figure.

To measure the dispersion of wages within these diverse groups we may construct what the statisticians call Lorenz or Gini curves, placing on the abscissae the cumulative percentages of wages and on the ordinates the cumulative percentages of recipients. Examination of these curves shows that the dispersion is absolutely identical for manual workers and for "cadres" of middle rank, that it is somewhat more unequal for clerical workers and much more so for "cadres" of the superior rank.

All in all, considering the actual facts, the salaries of "cadres" are sharply distinguishable from the wages of manual, or the salaries of clerical workers. For one thing, they are on a much higher level; for another, they are more widely dispersed; the category is less homogeneous, at least as concerns amounts disposable.

B. If we wish to proceed beyond the present state of affairs, and consider its evolution, we come to a very delicate question, one which arouses violent passions, that of determining whether the salaries of the "cadres" vary together with the wages of manual workers, or whether they have been falling relatively.

The only over-all statistics which are available are those drawn

Net Annual Wage or Salary, Full-time
Equivalents, Men and Women
(in thousand francs)

Index of Disparity
(wage of manual workers = 100)

Year	Manual Workers	Clerical Workers	"Cadres"		Manual Workers	Clerical Workers	"Cadres"	
			Middle Rank	Superior Rank			Middle Rank	Superior Rank
1947	97	120	175	319	100	124	180	328
1950	228	279	398	780	100	118	174	342
1951	262	319	550	994	100	122	209	379
1952	305	374	656	1200	100	122	215	393
1954	345	391	679	1408	100	114	197	408

from examination of the "Etats 1024." These statistics are very reliable, but unfortunately are not available prior to the year 1947.⁷

It appears that both the middle and superior ranks of the "cadres"

7. *Statistiques et Etudes Financières*, Supplément statistique, 1950, no. 5-6, pp. 181 ff.; *Bulletin mensuel de statistique*, suppléments, Oct.-Dec. 1952, p. 40; Oct.-Dec. 1953, p. 50. Oct.-Dec. 1954, p. 57; Oct.-Dec. 1955, p. 39; *Etudes Statistiques*, Jul-Sept. 1956, p. 39. The first examination of the "Etats 1024" was made for the year 1947, the second for the year 1950. A scrutiny for 1953 included only the provinces, to the exclusion of Paris.

have clearly improved their positions relative to the manual workers. The index of disparity for the middle rank "cadres" rises from 180 in 1947 to 197 in 1954, and the index for the superior rank "cadres" rises from 328 to 408. These figures support the thesis of the autonomy of the salaries of "cadres" relative to the wages of manual workers. However, the period covered is short and particularly troubled, an inflationary rise of prices having been pursued from 1947 to 1952.

When we shift to the period prior to 1947, our only data relate to particular groups of "cadres" which belong more commonly to the public or nationalized sectors than to the private sector.

a. As concerns the private sector, a table pertaining to metallurgy in the Paris region permits us to carry our study back to the year 1914.⁸ This table does not give us raw data, but only indexes calculated relative to the wages of manual labor, which correspond to our indexes of disparity.

Year	Manual Worker	Professional Worker	Engineer			Chief Engineer
			Beginner	After 2 yrs.	After 8 yrs.	
1914	100	188	200	300	400-600	900-1200
1920	100	150	138	170	244	
1921	100	168	187	210	350	480-600
1927	100	159				
1936	100	147	174	209	274	
1938	100	141	160	189	247	
1945	100	170	185	230	330	400-600
1946	100	143	155	193	316	
1947	100	142	153	184	291	
1948	100	140	150	181	282	
1951	100		200	240	360	400-800

The table indicates important variations in the situation of the "cadres" relative to that of workers. The index of disparity for the engineer with two years' service, for example, is 300 in 1914. It falls to 170 in 1920, rises to 210 in 1921, falls to 176 in 1943, rises to 230 in 1945, falls to 181 in 1948 and establishes itself finally at 240 in 1951. The index relative to engineers with eight years' service varies with even greater amplitude.

Representatives of the "cadres" often add that underlying these fluctuations is concealed a general movement of deterioration in their positions. This assertion requires discussion. That the relative situation of the "cadres" has fallen compared with 1914 is certain. But we must not lose sight of the fact that in 1914 the group was still

8. This table was drawn up by M. Vaillant, *Documents Jeunes Patrons* (May 1946) and *Liaisons Sociales* (April 1951).

largely in an embryonic state, and in any case very different from what it is today. Moreover, as Jacques Lecaillon has stressed,⁹ the figures prior to 1936 provide data from small samples and correspond to actual remunerations, while the later ones are minima fixed by the state or resulting from collective bargaining. Finally we should point out that practically all the deterioration had taken place by 1920. If we compare the figures for 1951 with those for 1920 we find not deterioration but improvement.

It would therefore be unwise to conclude from these figures, unique and somewhat uncertain, that there has been any decline in the relative position of the "cadres." We should remember only that this relative position varies, sometimes improving and sometimes worsening, while the general trend remains uncertain.

b. As concerns the nationalized sector, we have two established statistical series, one for employees of the tobacco monopoly and the other for the employees of the railways. The scope of these two

	Tobacco and Match. Monopolies ¹					National Railways Personnel ²		
Year	Annual Wage or Salary (in thousand francs)			Indexes of Disparity		Annual Wage or Salary (in thousand francs)		Index of Disparity
	Workmen	"Cadres"	Middle Superior	(Workmen = 100)	"Cadres"	Unskilled Worker	Division Engineer	(Unskilled Worker = 100)
1910						1.05	10	952
1930						12	61	508
1938	18.5	26	67.5	140	365			
1947						146	957	409
1952								
1955						264	668	254

1. Investigation by the "Service d'exploitation industrielle des Tabacs et Allumettes," reproduced by Marc Penouil, "Les Cadres et leur revenu" (mimeographed édition, p. 221; volume currently in press).

2. According to the "cadres" magazine, *Le Creuset* (Jan. 1, 1948 and Jul. 6, 1950) reproduced by Penouil, *ibid.*, p. 221.

inquiries is unfortunately very limited. In addition the data suffer from certain defects in the survey carried on by the "Service d'exploitation industrielle des Tabacs et Allumettes"; Penouil estimates "that the choice of bases for the calculations seem to present the situation in a manner systematically unfavorable for the personnel of the service." As for the inquiry regarding the railroads, one may inquire whether the same may not have been the case.

With these reservations these two inquiries show a net deteriora-

9. Jacques Lécaillon, "Le revenu des cadres," *Revue Economique*, 1952, p. 227.

tion in the situation of both middle and superior "cadres," not only between 1910 and the present time but also since 1930 or 1938.

c. As concerns the situation of the "cadres" of public officials, we have better information. For the period prior to World War I, the reports of the "Service de la Statistique générale de la France" date back to the year 1871.¹

Annual Salary Level (of officials 1911)	1871	1901	1906	1911
<i>Paris:</i>				
Over 25,000 francs	105	100	100	100
12,000 – 25,000	89	96	98	100
6,000 – 12,000	82	97	98	100
3,000 – 6,000	82	94	96	100
Less than 3,000 francs				
Agents	97	101	102	100
Subagents	67	85	93	100
<i>Outside Paris, or Unspecified:</i>				
12,000 – 25,000 francs	91	96	100	100
6,000 – 12,000	82	98	99	100
3,000 – 6,000	75	93	94	100
Less than 3,000 francs				
Agents	87	95	96	100
Subagents	69	87	93	100

If we assume somewhat arbitrarily that officials paid more than 3000 francs a year are "cadres" of middle and superior rank, while agents paid less than 3000 francs a year are clerks and subagents are workers, it appears that the remuneration of the "cadres" has clearly risen less than the remuneration of the workers, both in Paris and in the provinces. And none has risen less than that of the "cadres" at the upper range of the hierarchy.

In Paris, from 1871 to 1911, the inferior "cadres" gained 18 points, while the most elevated "cadres" lost 5 points and the workmen gained 33. The situation is the same in the provinces.

Regarding the period subsequent to World War I, we are much better informed, thanks to the studies of Tiano.²

1. *Rapport de la Statistique Générale de la France*, Annexe au Journal Officiel, Sept. 30, 1911. The table is reproduced in François Perroux, *Les traitements des fonctionnaires en France* (Paris, 1933), p. 139. See also Lucien March, "Contribution à la statistique des fonctionnaires," *Bulletin de la Statistique Générale de la France*, Oct. 1913.

The figures reproduced are indexes calculated using 1911 as the base year, or 100. They are not averages in the ordinary sense, but averages of the two extreme salaries in each category.

2. André Tiano, "Le traitement des fonctionnaires." (We cite the mimeographed edition. A printed edition is in press [Paris: Armand Colin].)

The civil service law of 1946 distinguishes four categories: "cadres" of superior rank (civil administrators, professors and assistant professors, engineers, etc.) who constitute Category A; "cadres" of middle rank (administrative secretaries, instructors, schoolmasters, etc.) who constitute Category B; white-collar workers (postmen, clerks, etc.) who constitute Category C; and purely routine workers who constitute Category D. Administrators and statisticians have found it convenient to define each category by the upper and lower limits of the salary scale of the civil servants which it includes. Category A therefore corresponds to index numbers from 225 to 800, Category B from 185 to 360, Category C from 130 to 250, and Category D from 100 to 185. We may suppose that Category A includes "cadres" of superior rank, Category B "cadres" of middle rank, Category C clerks, and Category D workers.

Tiano, considering first of all only gross remuneration, has followed the evolution of the "hierarchical fan" among the unmarried civil servants of the Paris region, situated at the index values 100 (workers), 130 and 185 (clerks) 300, 500, and 800 ("cadres"). He concludes that in 1956 these coefficients had diminished relative to their size in 1914 for all these officials: slightly for those with indexes of 130, noticeably for those with indexes of 185 and 300, and considerably for those with indexes 500 and 800.

Relative to 1930 the same divergences are found once more, but they are less marked. The relative position of officials with indexes of 130 had improved; that of officials with indexes of 185 had not changed; that of officials with indexes of 300 and 500 had fallen slightly, and that of officials with indexes of 800 had fallen back.

For the period beginning in 1938, a study of net remunerations shows that the action of the tax authorities had tended to reinforce the tendency toward narrowing of gaps within the hierarchy.

On the whole, if the income of the "cadres" is undeniably a wage since an essential part is a contractual remuneration granted in exchange for work, it does not seem to be a wage like other wages. We have successively pointed out that it is a monthly salary, that it often contains income in kind and a participation in profits, that it is generally kept secret by its recipient, that its average level is well above the average level of workingmen's wages or clerical salaries, that its coefficient of dispersion is higher, that it is subject to autonomous variations, and finally, at least as concerns the administrative and nationalized sectors, it seems to show an underlying tendency toward relative deterioration.

It seems then that we may anticipate without imprudence some

special behavior of the group called "cadres" with regard to the maintenance of its income, some breach in the unity of the labor market. In terms of the theory of distribution, the "cadres" constitute a fundamental type of participant in the distribution of the national income, a type which we should distinguish from others if we wish to arrive at a realistic description of the distributive process, or at a model sufficiently close to the world in which we live.

To what is this autonomy due? By what factors can it be explained? This is the final question which remains to be faced.

III

We need not insist on the fact that, to constitute a distinct category, the "cadres" must be *sufficiently numerous*, which assumes a *sufficient degree of accumulation of technical capital* and as a consequence sufficiently powerful enterprises and a sufficiently developed and organized state. Just as in French and English capitalism at the beginning of the nineteenth century, the "cadres" constitute, as we have shown, only an embryonic group of limited strength, there will be only very small technical "cadres" in enterprises which remain small in size and there will be only very small administrative "cadres" in a state which concerns itself only minimally with the economy.

Our problem is therefore only to understand the causes for the autonomy of the "cadres" in the modern French economy. We have shown above that the "cadres," while salaried employees, are characterized by both general and technical training surpassing the norm of their country. This being the case, we are faced with three questions: Have all citizens, from the moment they show adequate aptitudes, the possibility of acquiring the training required to penetrate the group called "cadres?" Supposing this not to be true in the present French situation, how is this to be explained? Finally, would eliminating the obstacles be enough to permit reabsorption of the category of the "cadres" within that of workers in general?

A. In France, we are told, education is gratuitous. The "cadres" therefore do not constitute a caste, but rather an open group. Let us see nevertheless whether, socially speaking, the "cadres" remain consolidated with other employees. To answer this question, we may make use of several recent studies.

a. A survey carried out by the Institut National d'Etudes Démographiques in 1950 and reaching more than 3,000 individuals permits us first of all more precise knowledge of the occupations of the fathers of members of the "cadres."³

3. Marcel Brésard, *op. cit.*, p. 539.

It appeared that 13 per cent of the "cadres" are sons of workers, 57 per cent sons of clerks, of subordinate functionaries, of artisans, or of working farmers, with 29 per cent — nearly a third — sons of members of "cadres," of businessmen, or of practitioners of liberal professions.

All other groups of salaried employees present very different proportions. One may well expect something of the sort with manual workers in both industry and agriculture. But if we concentrate our attention on clerical workers, who along with the "cadres" constitute the white-collar workers, we find that 30 per cent are sons of workers as against 13 per cent for the "cadres," and that only 6 per cent are sons of members of "cadres," of businessmen, or of practitioners of liberal professions, as against 29 per cent for the "cadres."

The group displaying proportions closest to those of the "cadres" is a group of *nonsalaried workers*, namely, businessmen and practitioners of liberal professions. Ten per cent of these are, in effect, sons of workers, 47 per cent sons of clerks and subordinate functionaries, and 43 per cent sons of members of "cadres" or businessmen.

b. Utilizing the same survey, we obtain similar results if we consider the social origins of wives, as indicated by the occupations of fathers-in-law.

The "cadres" contrast sharply with other wage earners, salaried employees as well as manual workers. They hardly ever marry the daughters of workers (6 per cent). More than a third of them (34 per cent) marry the daughters of members of "cadres," of businessmen, and of practitioners of liberal professions, *which other salaried employees practically never do*. Among the clerks and subordinate functionaries in particular, 31 per cent of the wives are daughters of workers, 5 per cent daughters of members of "cadres," of businessmen, or practitioners of liberal professions, and 63 per cent daughters of clerks, subordinate functionaries, artisans, or working farmers. For industrial workers the corresponding proportions are: 51 per cent, 1 per cent, 47 per cent, and for agricultural workers: 45 per cent, 1 per cent, 54 per cent.

This behavior of the "cadres" is very similar to that of businessmen and practitioners of the liberal professions who, more or less, choose their wives from the same groups: 10 per cent of business and professional men marry daughters of workers, 36 per cent daughters of members of the "cadres" or of businessmen, and 54 per cent daughters of clerks and subordinate functionaries.

c. From the occupations of fathers and fathers-in-law we may

pass to scales of living, using another inquiry of the Institut National d'Etudes Démographiques relating to the year 1948.⁴

The survey concluded that more than 25 per cent of the "cadres" own an automobile; while only 16 per cent of the subordinate functionaries, 5 per cent of the clerks, and 2 per cent of the workers of all categories own one.

More than 20 per cent of the "cadres" employ a domestic servant, while only 12 per cent of the subordinate functionaries, 2 per cent of the clerks, and 1 per cent of the workers employ one.

More than 30 per cent of the "cadres" have a telephone, while only 9 per cent of the subordinate functionaries, 6 per cent of the clerks, and 1 per cent of the workers have one.

Do the "cadres" approach the businessmen and practitioners of the liberal professions? Certainly, but less so than one might suppose from studying the occupations of their fathers and the social origins of their wives. A definitely higher proportion of businessmen and of practitioners of liberal professions own an automobile (82 per cent and 56 per cent against 25 per cent), employ a domestic servant (67 per cent and 46 per cent, against 20 per cent) and have a telephone (87 per cent and 75 per cent, against 30 per cent). Sons therefore pass from one group to another, and there is intermarriage, but the businessmen and professional practitioners have a scale of living superior to the "cadres."

d. A final inquiry made by Jacquin⁵ and pertaining to 289 engineers of the Ecole des Arts et Métiers permits us to complete the picture.

This inquiry shows that the "cadres" devote to purchases of durable goods or to increasing their liquid reserves sums easily in excess of 20 per cent of their incomes. As a result of these investments or in consequence of inheritances, Jacquin concludes that 30 per cent of the "cadres" possess incomes other than their salaries, incomes amounting to 10 per cent of their total resources. Because of the reticence of the interested parties, we may suppose that these figures are somewhat understated. "We may conceive," concludes Jacquin, "the well-being and above all the considerable degree of security which sets off such a type of living from that of the manual or clerical proletariat."

In sum it appears that though the "cadres" are undeniably wage earners, they are not to be confused with other wage earners. Not only are they distinct from manual workers, but within the white-

4. *Ibid.*, p. 541.

5. François Jacquin, *op. cit.*, p. 141.

collar group, they are equally distinct from clerical workers. Their social origins and their manner of living are both different.

B. It is always a problem to determine how the "cadres" can safeguard this autonomy which we have just established, vis-à-vis other wage earners and particularly vis-à-vis clerical workers. This is due in large measure, we believe, to the character of the process of capitalization since, as applied to labor, it gives rise to the special commodities, the services of the "cadres."

Within the social groups from which the "cadres" come, the expenses of children's education are generally considered as *consumption expenditures* which the head of the family feels himself morally obligated to include in his budget with a priority exceeding many other expenditures, particularly the raising of his own scale of living. A certain proportion of children having gained by these expenditures, arrive in this way at the category of the "cadres." These individuals are limited in number. They bring with them certain concepts and certain manners of living, certain requirements, a certain evaluation of what a member of the "cadres" in French society might normally aspire to. As functions of their number and these requirements, levels of remuneration are set which are superior to those of workers and clerks and which need not vary parallel with the latter.

The excess of salary achieved by the "cadres" generally exceeds the *amount necessary to amortize the sums spent on their training*. Why does this fact not produce an influx of people from other social classes than those which normally furnish the "cadres"? Why, in these other classes, do parents not borrow the sums necessary to support children with the requisite aptitudes during their periods of training, since the amortization of these sums is not only possible but leaves a net gain? The answer may be found in the *special character of investments in individuals*.

Those who cannot treat the expenses of training their children as consumption expenditures, but must consider them as profit-making investments, encounter a double difficulty. On the one hand, they generally lack knowledge and cannot follow *that perfect long-period rationality* which theoretical economists assume in so facile a manner. On the other, one must give guarantees for loans, and the prospect of the ultimate success of one's children is not generally considered sufficient. As a result only a minority of children coming from these classes can attain the status of "cadres." They conform to the behavior of the other "cadres," but their rise does not affect the autonomy of the group.

C. Would this state of things be modified by a wider diffusion of instruction and of public investments complementary to inadequate private investments, as regards training and support during the period of training? Might we consider, in other words, that the natural distinction existing between the behavior of the "cadres" and that of the lower strata of wage earners is derived from the structure of the French economy, and that it might disappear with sufficient modification of this structure? Or must we admit that the same process of capitalization which has resulted in the formation of the "cadres" presents characteristics such that always, whatever the economic structure may be, a threshold will separate the behavior of the "cadres" from those of the lower strata of wage earners?

To answer this question it would be necessary to possess information on what goes on in countries with different economic structures, such as the United States or, if we wished to maximize the differences, the Soviet Union. We do not know whether such a study has been made for the United States. As regards the Soviet Union, we lack information on which to judge whether, in fact — economic doctrine makes no difference here — the managerial "cadres" constitute a category apart from the lower strata of wage earners in relation to the distribution of the national income. We nevertheless suggest, subject to later verification, that even if social mobility prevents to some extent the crystallization of separate group behavior patterns, it does not prevent the majority of individuals being confined throughout their lives in the same occupational blind alleys. It seems difficult to unify the conditions of work of the "cadres" and of manual or even of clerical laborers to such a point that they become interchangeable. In consequence our distinction between their two categories of income will never entirely disappear.

In sum it appears to us that, in the present structure of the French economy, one must isolate the income of the "cadres" from the income of other workers in order to understand the problems of the distribution of the national income. They are two different types of wages. Is this perhaps likewise true in other economies?

JEAN MARCHAL.

UNIVERSITY OF PARIS

JACQUES LÉCAILLON.

UNIVERSITY OF LILLE

INTERNATIONAL SPECIALIZATION AND THE CONCEPT OF BALANCED GROWTH*

By JOHN SHEAHAN

Introduction, 183. — I. The problem of investment incentives in a low income country, 184. — II. Internal difficulties in the argument for balanced expansion, 188. — III. The role of international trade, 191. — IV. Interdependence of the rate of return on capital and the rate of economic growth, 194. — V. Conclusions, 196.

"Balanced growth" is one of those happy terms that seduce the mind into acceptance rather than analysis of the idea intended. The view explained in this paper is that the whole concept is an intriguing monstrosity: it contains a very useful thought, submerged under a number of potentially quite costly confusions. The valid element traces back to Allyn Young and Rosenstein-Rodan. It was clearly spelled out by Scitovsky and presented in most persuasive terms by Ragnar Nurkse.¹ But in Nurkse's exposition the lurking difficulty in the argument moved fully onstage, embodied in that mischievously mystical term "balance." It remained for Arthur Lewis to take the courageous step of providing a fairly definite meaning for the term, thereby turning the argument into an attack on the principle of international specialization.²

As it stands, the discussion contains so many useful and appealing points that the misleading elements are not properly recognized. The importance of the possible errors in interpretation is considerable, because the argument easily lends itself to exactly the wrong orientation of investment policy. The idea of balance needs to be taken apart, so that the confusing strands may be separated from the genuine underlying insight. The present paper is an attempt to do this, focusing the inquiry on Nurkse's explanation of the concept.

* Paul G. Clark and Emile Despres, RAND Corporation and Williams College, aided greatly in exploration of several of the questions basic to this paper.

1. Young, "Increasing Returns and Economic Progress," *Economic Journal*, XXXVIII (1928), 527-42; Rosenstein-Rodan, "Problems of Industrialization of Eastern and South-Eastern Europe," *Economic Journal*, LIII (1943), 202-11; Scitovsky, "Two Concepts of External Economies," *Journal of Political Economy*, LXII (1954), 143-51; Nurkse, *Problems of Capital Formation in Underdeveloped Countries* (New York, 1955), chap. 1.

2. Lewis, *The Theory of Economic Growth* (Homewood, Illinois, 1955), pp. 276-83.

Marcus Fleming's careful article on the same subject, and Nurkse's reply, are also used to help clarify the basic assumptions.³

The discussion here is essentially concerned with demonstrating that: (a) the concept of balance may be usefully interpreted to support co-ordination of investment decisions, based on the principle of interdependence between the rate of return on capital and the rate of economic growth, but (b) it does not include any valid point supporting a presumption in favor of deliberate diversification. If "balanced growth" includes the case of concentrated development of one or a few industries — i.e., of marked changes in the composition of output, not necessarily co-ordinated with the pattern of domestic demand—then the point is acceptable. If it is meant to support a policy of matching production to demand changes by expanding most or all industries simultaneously, the position is untenable. In any case, a direct consideration of the implications for policy may serve to reduce the scope for unhelpful inference from Nurkse's creatively ambiguous generalization.

I. THE PROBLEM OF INVESTMENT INCENTIVES IN A LOW INCOME COUNTRY

One of the distinctive elements in Nurkse's stimulating discussion of economic growth problems was his emphasis on the role of insufficient demand as a factor inhibiting investment in low income countries. His concern was not with aggregate monetary demand in the familiar Keynesian sense, but with the effect on investment of a prevailing low level of income. "It is not a deficiency of 'effective demand' in terms of Keynesian economics. There is, as a rule, no deficiency of monetary demand, there is no deflationary gap." Rather, "on the demand side, the inducement to invest may be low because of the small buying power of the people, which is due to their small real income, which again is due to low productivity."⁴

Traditional analysis does not provide an established home for the proposition that a low level of real income acts to inhibit investment on the demand side. It suggests the opposite: in so far as low income is associated with a low ratio of capital to co-operating factors of production, the marginal product of capital should be higher than in capital-rich countries, and investment should pay rather well. This

3. Fleming, "External Economies and the Doctrine of Balanced Growth," *Economic Journal*, LXV (June 1955), 241-56; Nurkse, "Balanced Growth on Static Assumptions," *Economic Journal*, LXVI (June 1956), 365-67; Fleming, "Rejoinder," *Economic Journal*, LXVI (Sept. 1956), 537-39.

4. Nurkse, *Problems of Capital Formation in Underdeveloped Countries*, p. 17, p. 5.

view is extremely simplified but is not necessarily invalidated by the low ratio of investment to national income actually observed in poorer countries. The latter can be explained, perhaps too readily, by such factors as a social-cultural environment inhibiting entrepreneurship, or by the assumption that a low level of income necessarily results in strong pressures to consume all of current output. Neither of these suggestions can or should be ruled out; the latter is very often likely to be the dominant difficulty. But Nurkse's point cuts deeper analytically; is it possible, even given a climate of rational calculation by competent entrepreneurs with resources available for investment, that the process of capital accumulation will be systematically thwarted because individual investment projects cannot pay? If so, if investment must proceed in a balanced pattern meeting all directions of demand simultaneously, then a number of basic arguments from factor proportions, and traditional ideas of specialization and trade, need serious revision.

In order to focus on the "balance" concept at issue, it is desirable to rule out of consideration some very real problems. It is assumed here and throughout this paper that, in the countries considered: (a) there are intelligent businessmen interested in possible investment if they can foresee a reasonable minimum rate of return, (b) that the potential entrepreneurs have high incomes or other means of access to resources which would make it possible for them to carry out investment, (c) that the minimum essential "social overhead capital" is being provided by the state, and (d) that we have no problem of deficient aggregate monetary demand. In order to define (d) more clearly, assume Say's Law: if the would-be entrepreneurs do not find any rewarding investment possibilities, they spend all their income on consumption. I.e., total investment varies within some moderately wide range depending on the attractiveness of possible projects, and intended saving always adjusts to match intended investment.

Nurkse's special argument is that, for any *single* investment project, the poverty of the people will make it extremely difficult to sell the additional product made possible by investment. In his illustrative case, "Where the great majority of the people are too poor to wear leather shoes, setting up a modern shoe factory may be a doubtful business proposition."⁵ The trouble is stated to be caused by the fact that the workers in the shoe factory will want to spend most of their income on other goods, and the rest of the people will not readily deflect their buying from old essentials to the new products.

5. *Ibid.*, p. 7.

"The trouble is due by no means solely to discontinuities in the technical forms of capital equipment, though these will accentuate it. It is due above all to the inevitable inelasticity of demands at low real-income levels . . . The difficulty caused by the small size of the market relates to individual investment incentives in any single line of production taken by itself. At least in principle, the difficulty vanishes in the case of a more or less synchronized application of capital to a wide range of industries . . . People working with more and better tools in a number of complementary products become each other's customers."⁶

The argument is troublesome. On the surface, Nurkse has falsely fortified his case by a misleading example. Somewhat more deeply, there is a true analytical problem. To clarify the surface level difficulty, consider the example of the shoe factory as a possible investment in a very poor country, where people spend nearly all their income on food and simple textiles, and would spend most of any additions to income on such essentials. By definition, we have set up a situation in which the demand for shoes is inelastic. The price at which all the output from a shoe factory of minimum size can be sold may very well be so low that the investor will lose if he goes ahead, and will not go ahead if his foresight is good. But it is not correct to conclude that a worthwhile investment opportunity will be foregone because of any inevitable inelasticity of demands." The particular investment envisaged *should not* be undertaken. If the price at which the shoes may be sold will not cover costs of production, the presumption must be that people do not want them, at the price of the other things they would have to give up to buy the shoes. If we do rule out of consideration the possibility of deficient aggregate demand, then money incomes earned in shoe production must be going somewhere. The trick would seem to be to invest in production of goods people want at the level of income they will have when the investment project is carried out.

But there is a genuine analytical problem which may best be brought out by retreating from realistic examples into algebraic notation and by specifying a simplified set of initial conditions. Consider two commodities only, glamorously tagged x_1 and x_2 .

Let p_1 and p_2 denote their prices

u_1 and u_2 denote their marginal utilities.

And let R denote the minimum acceptable rate of return on capital at which investors will choose capital formation over consumption.

6. *Ibid.*, pp. 9-11.

Assume that, initially,

$$(1) \quad \frac{u_1}{p_1} = \frac{u_2}{p_2}$$

- (2) The rate of return on capital in each industry is just equal to R .
- (3) Potential investors interested in each industry are considering projects which could secure discrete increases in production at unchanged unit costs.

Taking separately the industry producing x_1 , expansion of supply will increase aggregate income and raise demand for both commodities. But if the supply of x_2 is unchanged, u_1/u_2 will decrease.⁷ P_1 will then decrease to restore equality between u_1/p_1 and u_2/p_2 . If p_1 does decrease, then the rate of return on capital goes below R , the minimum postulated as acceptable to investors. The potential investor, if gifted with foresight, will reject investment in production of x_1 .

The objection made above concerning Nurkse's example was that the case seemed to rest on the simple assumption that x_1 was a "non-essential" for which the elasticity of demand was low by definition. But clearly the problem does not depend on any particular demand elasticities: it applies to expansion of either commodity individually, supply of the other commodity remaining constant.

Given the specified initial conditions, expansion by either industry alone would bring the rate of return in that industry below R . But it is clearly possible that a joint expansion, increasing the supply of both commodities, could eliminate the difficulty. The requirement would be that both industries expand production at exactly the rates consistent with the income elasticities of demand for the two commodities, maintaining $u_1/p_1 = u_2/p_2$ at constant prices. This seems to be close to the meaning of "balance" specified by Lewis: not that relative outputs remain constant, but that they change exactly in line with relative demand at rising income levels.⁸

If the concept of balanced growth is intended to refer to cases similar to the one specified here, it does have a logical basis. The argument need not be tied down to such a special case. The general possibility of insufficient investment in cases where income could be increased by joint expansion remains open if we relax the assumption

7. The change in relative marginal utilities and prices would, of course, be more marked if the expansion of x_1 required a decrease in the production of x_2 . This possibility will be discussed below.

8. Lewis, *op. cit.*, p. 278.

of constant costs in both industries.⁹ But once relative prices are allowed to change, the concept of balance becomes elusive again. In particular, income elasticity of demand for any commodity loses specific meaning.¹ It will best bring out the possible sources of trouble in the argument if we proceed by progressively relaxing the suggested initial conditions in order to examine difficulties in this unduly rigid but relatively definite model.

II. INTERNAL DIFFICULTIES IN THE ARGUMENT FOR BALANCED EXPANSION

The essence of the preceding suggested meaning for the concept of balanced growth is that investment should be carried out in diverse industries as indicated by the pattern of income elasticities of demand for all products. The argument presents at least three overly inviting paths toward misinterpretation. The following possible statements should be rejected: (1) individual investment projects will not be profitable in the absence of balanced expansion, (2) if investment is carried out in such directions as to match production with all increases in demand that would occur at constant prices, the rate of return on capital will remain constant at an adequate level in all industries, and (3) since domestic expansion raises domestic but not foreign demand, investment should be guided by the pattern of domestic demand only. The third point will be examined separately in Section III following.

The first of the three possible errors is trivial in the sense that no one would defend it, but less trivial in the sense that explicit recognition of what is involved leads to the conclusion that the problem of weak investment incentives may well be nonexistent in many cases. The argument that x_1 will yield a return below R if output of x_2 remains constant rests on the assumptions that: (a) the initial situation is one of equilibrium in which returns on capital in each line of production are just equal to R , and (b) technological changes in the form of opportunities for lower cost production or product innovation are excluded. If these conditions were applicable, there would not be much investment in any country, regardless of the level of income. With respect to (a), if the actual rate of return on capital is $3R$, then the fact that a new plant may reduce p_1 and lower the rate of return need not discourage investment.² With respect to (b), if new invest-

9. See especially the lucid alternative formulation by Scitovsky, *op. cit.*, pp. 145-51.

1. It is not a clear concept at best. Cf. P. DeWolff, "Income Elasticity of Demand, a Micro-Economic and a Macro-Economic Interpretation," *Economic Journal*, LI (1941), 140-45.

2. This situation might well discourage investment by firms already in the

ment can lower production costs for x_1 , then a decrease in p_1 will not necessarily lower the rate of return. Finally, where the project would make possible the introduction of a new product, there is surely no necessary presumption that its price will be such as to yield a return below R .

The balanced growth argument thus should not be understood as a general proposition that, in a country with a low level of real income, demand problems will "inevitably" foreclose individual projects. In fact, the separate projects that are capable of yielding a return above R may at times collectively exceed the possible capital formation of the country, even given a fairly elastic supply of capital. Many underdeveloped countries seem to have little or no difficulty with deficient investment demand.

The second suggestion to be examined is that joint expansion could maintain the rate of return on capital at R in all industries, even assuming that all returns are initially at this level and that technological change is excluded. Again, there is a high possibility of confusion. Consider all practical problems to be solved: assume that the rate of income growth and pattern of demand at changing income levels can be accurately predicted, and that the government can assure all firms that prices of their products will remain unchanged if they expand supply at properly stipulated rates. The special conditions required to make such a program work are: (a) a highly elastic supply of all factors of production, (b) unchanging relative costs of all goods, (c) perfect divisibility for all factors.

Assuming perfect divisibility for the moment, consider the possibility of inelasticity in factor supplies. If production of x_1 increases, it may raise factor costs in production of x_2 so steeply as to force contraction rather than encourage joint expansion.³ The balanced growth concept must assume either excess supplies of productive factors other than capital, or highly elastic supplies if the situation approximates equilibrium. Such assumptions do seem valid for many underdeveloped countries. But they should not be pushed to the point of postulating a stable structure of relative costs. Expansion on a wide front is almost certain to run into specific input limitations, field, but not new entry. Policies favoring free entry and competition are particularly vital for underdeveloped countries starting from a position in which there are few firms in any industry. Cf. the more general argument of Henry Bruton, "Innovations and Equilibrium Growth," *Economic Journal*, LXVI (Sept. 1956), p. 466.

3. Fleming, *op. cit.*, p. 246: "The situation might be roughly expressed by saying that, whereas the balanced-growth doctrine assumes that the relationship between industries is for the most part complementary, the limitation of factor supplies ensures that the relationship is for the most part competitive."

forcing changes in the pattern of relative costs. In any such case, production should be "unbalanced" in the interest of maximizing income growth — i.e., concentrated in directions of falling relative costs rather than diversified to match the pattern of increases in demand at constant prices.

To see the implications of changes in relative costs, let the cost of producing x_1 rise while the cost of x_2 remains constant. If expansion is to be completely balanced in the sense of matching all increases in demand that would take place at constant prices, investment in x_1 will yield a return below R . Under market conditions, production of x_1 would not be increased, and p_1/p_2 would rise. If the ideal of balance were taken to extreme, it would be possible to keep relative prices constant by subsidizing investment in x_1 . To do so would lower the possible growth of real income: u_1/u_2 is constant by hypothesis, but increasing amounts of x_2 must be given up per unit of x_1 secured (because the latter requires increasing factor inputs per unit of output). Conversely, if x_1 could be produced at decreasing cost as scale rises, market reactions would tend to increase x_1 faster than demand: the economy would become, and should become, "unbalanced" in the sense that p_1/p_2 would decline.

Any ideal of balance in terms of demand changes alone must therefore be rejected as an unnecessarily costly objective. It would neglect one half of the picture — that of the limitations and possibilities opened up by changes in relative costs. When this point is admitted, it becomes quite difficult to specify any meaning for "balance" that is both exact and interesting. In fact, it forces a return to the ancient principle that investment should be so directed as to keep returns approximately equal in all industries. The latter could mean diversified expansion if relative costs remained constant; it could mean complete specialization on one product if its cost fell more rapidly relative to other costs than its price fell relative to other prices. If the nation's investment were being planned by one central decision unit, with the objective of raising real income as rapidly as possible, it should not be guided by any predisposition toward diversification. If investment is left to multiple independent decisions, there is no reason to try to promote joint expansion by all private investors.

Apart from the problem of changes in relative costs, the question of indivisibilities in factors of production remains to be examined. Nurkse is correct that this is an important consideration with respect to capital equipment. His view, quoted above, is that such indivisibilities add to the individual investment demand problem and, by

implication, re-enforce the argument for balanced growth. The logical deduction is exactly the contrary.

The basic problem of individual investment incentive is that expansion of x_1 must be expected to lead to a decline in its relative price. The larger the minimum scale on which investment can be carried out, the greater the problem. But it does not follow that, if the situation is indeed one in which each individual project must be carried out on a large scale, it would make sense to try to carry out large projects in all directions at once. Given any upper limit on the resources available for investment, a balanced program of simultaneous large scale investment in all fields becomes an impossibility. The real problem of capital indivisibilities accentuates the familiar principle of selecting the few best possibilities. The importance of selectivity gains rather than loses relevance in a low income country.

The immediately preceding argument seems to lead to an utter impasse: if large scale investment projects are required, they do accentuate the problem of the rate of return on each individual project, but they surely also reduce any possibility of securing balanced growth. The key to a solution is to reject the principle of domestic balance, which amounts to tying an unnecessary anchor to an economy, and take advantage of the opportunities provided by trade.

III. THE ROLE OF INTERNATIONAL TRADE

The reason that the individual investment demand problem arises is that the expansion of supply of any one commodity, all others being constant, will normally lead to a decrease in the relative price of the good increased in supply. The case for balanced growth starts from the possibility that a joint expansion in all sectors will alleviate this problem of decreasing prices for each individual investment calculation. This case may be misleading in a most costly way because of its tendency to focus attention on production for domestic markets only (alternatively expressed, to plan to meet all demand increases from domestic output). This was the third of the possible interpretive errors suggested at the beginning of Section II above: since increases in domestic income raise domestic but not export demand, it may be assumed that investment should be guided entirely by the pattern of increases in domestic demand.

It is true that economic expansion in an underdeveloped country will not normally have a significant effect on demand for its exports, but this surely does not mean that all gains in output should be

directed to the home market to avoid falling prices. The possibility that supply expansion will have a negative effect on the price of the commodity concerned is very clear when we consider a closed low-income economy, but not when attention is shifted to the world market. For the latter, the increase in supply resulting from any individual project will normally be trivial. The extreme presumption, utilized by Samuelson in his classic proof of the gains from trade,⁴ is that external prices will be completely unaffected by changes in demand and supply in any one country. If this assumption did approximate reality, and if our underdeveloped country were to achieve a domestic price structure permitting a balance between exports and imports, then the individual problem as posed by Nurkse would not exist.

The degree of unreality in the assumption that world prices will not be affected by expansion of output in any one country cannot be resolved in general terms. The normal effect must be that an increase in supply of any one commodity will decrease its relative price. But if we are considering products that can profitably be exported at current prices, it must surely be rather rare that the increase in supply from any one investment project will significantly alter the world price. Nurkse is quite right that large-scale expansion of agricultural output would have a negative effect on export prices in many cases, but this is not the investment incentive problem with which we are concerned. For individual projects in the industrial sector, few underdeveloped countries are likely to rock the world price structure. If the commodity involved can be exported at a profit, the demand problem for the individual investment project is largely mythical.

The more important difficulty concerns costs: given that the increase in supply from a single project will not normally change the world price for the commodity significantly, the producer in the underdeveloped country may often find it difficult to expand output without running into increasing costs or absolute input limitations. This problem will weaken investment incentives differentially, discouraging expansion of those goods for which costs increase, and not hampering investment in those which can be produced at constant or decreasing costs. Considering the latter lines of production only, we return to the conclusion that the investment incentive problem is partly mythical, if the proper policy of "unbalancing" production and developing only the country's relatively efficient industries is chosen.

4. "The Gain From International Trade." Reprinted in *Readings in the Theory of International Trade*, pp. 244-45.

But it would be highly unrealistic to consider the problems of investment incentives to be easily solved by a policy of promoting specialization and trade. Transport and market entry costs, effectively supplemented by deliberate import restrictions, often preclude access to external markets. More generally, the relative inflexibility of management techniques and the slow process of developing diversified skills in the underdeveloped countries make it highly difficult to keep adjusting effectively to world market conditions.⁵ It is also true in some cases that national economic policy aggravates the problem by maintenance of an overvalued currency, which is a supremely potent device for limiting the market. Currency overvaluation sustains conditions of market isolation in which the analysis of a closed economy becomes relevant. But the difficulty cannot be wished away. If there is a continuing one-way trend towards external deficits, arising from relatively low productive flexibility in the underdeveloped country, then the additional problems generated by repeated devaluation might well be too great to allow continuing adjustment of internal to external prices.⁶ I.e., we must expect national markets in underdeveloped countries to be cut off from world markets in some degree, even given complete avoidance of any deliberate policy of seeking domestic balance. To the extent that domestic prices do systematically exceed the domestic currency equivalent of possible export prices, the investment incentive problem remains a real consideration.

The idea of efficient resource allocation through trade is applicable with special force to the situation of an underdeveloped country. It encourages investment in those industries which can contribute most to raising income, and it avoids the high cost of striving for internal balance in an economy in which factors or production are just beginning to be oriented toward the flexibility required for diversified production. There is no genuine economic case for a policy of domestic balance. But there is a core of reality to the problem of investment incentives within limited markets where new projects add significantly to existing supply. This problem provides some basis for a policy of co-ordinating investment, once the issue is separated clearly from any misleading argument for balance between domestic production and domestic demand.

5. Cf. C. P. Kindleberger, *The Terms of Trade* (New York, 1956), chap. 10, esp. p. 253.

6. No one has better expressed the difficulties involved in a continuing process of devaluation than Nurkse, in *International Currency Experience*, League of Nations, 1944, chap. 5.

IV. INTERDEPENDENCE OF THE RATE OF RETURN ON CAPITAL AND THE RATE OF ECONOMIC GROWTH

The valid element in the balanced growth concept is the principle that the rate of return on capital is a positive function of the rate of economic growth. For an increment of capital of given physical productivity, the rate of return rises with the price of the product, and the price of the product rises with the supply of all other products. This relationship provides a possible basis for a policy of co-ordinated joint investment, though its usefulness is limited by several offsetting considerations.

To amplify the positive case for a co-ordinated investment program, consider first the operation of a unified central authority directed to formulate a plan to carry out all projects which would yield a return of, say, 20 per cent. In principle, the authority could take account of all the changes in factor and product prices that would result from any proposed program. Expansion of supply for any given product would tend to lower its relative price, but the offsetting effect of increasing supplies of other products could be anticipated. In contrast to this, if investment were determined in multiple private decisions taken independently, the potential investors could not rely on the helpful effect of increasing supplies of products other than their own. Except for export goods, the expected price of the product would (probably) be lower than the price anticipated by a central authority able to foresee all other supply changes. The individual investor's estimate of revenue may thus be said to be systematically biased downward: for the same target rate of return, fewer projects would appear to qualify.

One immediately apparent offsetting consideration concerns input costs. The individual investor may neglect favorable demand effects of simultaneous expansion in other lines, but if so he probably would also neglect unfavorable effects on the prices of required factors of production. A central authority would be able to foresee when two promising projects would each need more than half of the limited supply of a strategic input, and could accordingly reject one of them; two independent investors might jointly underestimate their costs and arrive at more optimistic expectations than would the central authority. This does not mean that the unco-ordinated approach is to be preferred — it could lead to very wasteful capital formation where co-operating factors are in short supply — but it does mean that the argument for a systematic downward bias in expected returns is weakened.

The degree to which this cost aspect is important depends on the elasticity of supply of all factors of production.⁷ If the particular country has abundant raw materials, and excess labor possessing required skills, the objection drops out of the picture and leaves a strong case for a co-ordinated program. The degree to which any particular country approaches these conditions is an empirical question to be investigated in each specific case.

The most important supply question concerns the availability of skilled labor and effective management. If these are in fixed supply, then (a) the case for joint capital formation is very weak, and (b) the country is not likely to get anywhere very fast whatever it does. Diminishing returns to capital would be a potent consideration: any one project would be likely to lower possible returns for others. But where countries are starting from disequilibrium conditions with excess manpower, and where strong efforts are being devoted to train required skills, the borderline of diminishing returns to capital may well be moving away faster than the supply of capital increases.

A further possible objection to the positive case for co-ordinated expansion is that independent investors are themselves capable of basing their calculations on the expectation of rising income, and thus need not consistently underestimate the probable prices of their products when supply is increased. This possibility must surely depend upon actual experience. If income has risen very slowly at best in the past, private entrepreneurs are likely to forecast investment results on the expectation of little future income change, and thereby to anticipate serious adverse effects on prices from discrete increases in supply. Where the aim is to replace past stability by rising income, a co-ordinated program may be necessary until the evidence of success begins to affect independent investment calculations.

The preceding arguments can be expressed in terms of the general concept of market interdependence. The demand for x_1 is a positive function of the supply of all other commodities. But an increase in the supply of other commodities *may* also raise the cost of production of x_1 ; it will not necessarily raise the rate of return on investment in its production. This negative cost effect will not result if productive factors are highly elastic in supply, or if technological change permits the expansion without raising input requirements. The presumption should ordinarily be that capital formation will, by increasing the total supply of productive factors in the economy, raise real income

7. Cf. Fleming's thorough discussion of cost aspects, note 3 above, p. 184.

and have a net positive effect on many, though surely not all, other industries.

The proposition here differs from Nurkse's argument in several respects, as amplified in this and preceding sections. The major difference here is the point that any co-ordinated program should be concentrated on those focal industries in which returns are highest, taking into account conflicting claims for scarce resources. Concentration on the few best industries should be the guiding rule of any program to raise income. The positive effect is completely disassociated from any concept of desirable diversification. The expected return on investment in production of x_1 is a function of the *supply* of all other products, not of the rate at which they can be produced domestically. If x_2 can be secured at a lower cost through trade (in terms of x_1 exported) than it can be obtained by diverting domestic resources from the production of x_1 , then income growth will be maximized by concentration on the production of x_1 . I.e., the incentive to invest is *not* a function of the degree of balance in domestic production, but of the growth of real income.

V. CONCLUSIONS

The appeal of the balanced growth concept is highly understandable. "Balance" just cannot be undesirable. But a policy based on a misleading interpretation of what economic balance means can be decidedly harmful. The literature on this concept lends itself much too readily to misinterpretation. The valid general point is sufficiently important to enlist interest, and sufficiently imprecise to invite trouble.

The genuine insight in the argument derives from a basic principle of general equilibrium. As explained in Section IV above, it is correct to state that the expected rate of return on investment in any specific project is a positive function of the rate of growth of real income. This relationship offers scope in some cases for raising the ratio of investment to income in the market-determined sector of the economy by a planned investment program which itself increases income. But this valid point can easily lead astray in two critical respects.

In the first place, the possible negative effect of capital formation on costs must be carefully assessed. This may be a dominant adverse consideration when supplies of co-operating factors are fixed and external economies not present in the particular projects undertaken. The strength of the positive case rests on the availability of co-operating factors in excess, or at least highly elastic, supply.

Secondly, the concept of balance may be very misleading for policy in so far as it suggests deliberate guidance of investment in directions matching the pattern of increases in domestic demand. It shifts attention away from the possibility of raising income more rapidly by concentrating on production of those goods which fall in relative cost, and trading them for goods which rise in relative cost as domestic expansion proceeds. Balancing domestic production and demand would make sense only in the most exceptional case that relative costs remained constant as output increased. Taking the real problem of capital indivisibilities into account further strengthens, not weakens, the case for specialization and trade.

There is a case for a co-ordinated investment program. It should normally include several industries, and should probably not in the usual case be concerned primarily with investment in agriculture. But the general objective of any such program, and guiding policy for the underdeveloped country, should be the traditional one of economizing in the use of resources by "unbalancing" production in the direction of goods produced with relatively greatest efficiency. There is no valid point in the concept of balanced growth which suggests any useful departure from this principle. The possibilities of gain from trade are, unhappily, limited by numerous real impediments. But the underdeveloped countries can hardly gain by independent policies multiplying such impediments. The return on investment projects in all of them will increase with the extent of the market which the investment may serve, and that extent is effectively decreased by pursuit of domestic balance. The term "balance" can be downright unhelpful applied to any one nation taken alone, but could have a most useful meaning as a policy of co-ordinating investment internationally.

JOHN SHEAHAN.

WILLIAMS COLLEGE

WHAT IS THE ECONOMIC SYSTEM?

By DAVID McCORD WRIGHT

Introduction, 198. — I. Concepts of equilibrium and the "machine," 198. — II. Difficulties with the concept of the "machine," 200. — III. The "event" approach, 202. — IV. Major limitation of the "event" approach, 205. — V. Implications of the "event" approach, 208. — VI. Summary, 210.

What *is* the economic system? In what terms may we best think of it? Of course there are basic objective data which remain the same no matter how we think of them, or, indeed, whether we think of them at all. But one of the complicated points about economics is that our thoughts and beliefs (and changes in them) are part of the "objective" data. And even if this were not the case there would still be few students of scientific methodology who would deny the importance, for good or ill, of the accuracy and usefulness of the basic analogies and abstractive schemes which any science employs. And so I ask again — what "*is*" the economic system — what set of concepts, what "picture," can we hold in our minds as being the most nearly accurate and the most useful description available?

I

Economics has made use of many analogies and many figures in its time. From about 1880, and until quite recently, for example, the implicit concept most employed was that of an "equilibrium." A number of forces *simultaneously* existing are thought of as pushing against each other, within a market, toward a state of balance and, once attained, this balance will remain undisturbed until some new force intrudes. The concept is, in a sense, "spatial" and not temporal. It focuses upon "adjustment" within an industry or area, not movement in time, or general expansion.¹ Thus the economy may, for example, be visualized as a bowl of water. Somebody drops in a brick. The water quivers. Then a new calm ensues.

But with the study of business cycles, and particularly with the advent of Keynesian economics, basic analogies have greatly changed — though the process has been, to a considerable extent an uncon-

1. The movement toward adjustment taking place over time, in other words, is disregarded since attention is focused upon the pre-destinate ultimate result determined by the initial "forces."

sciousness one. No longer are we thinking of balls in a bowl, or water in a bowl, or "forces" simultaneously pulling (or pushing) each other. We begin instead to think of the economic system in terms of the construction of, and flow of material through a sort of *machine*. There is supposed to be some great maximum "engine" that *could* be built which, taken in conjunction with the assumed labor force, tastes, resources, technique and so on *would* give the maximum possible standard of living. The aim is to get that machine built, with a minimum of waste and maximum of efficiency in the co-ordination

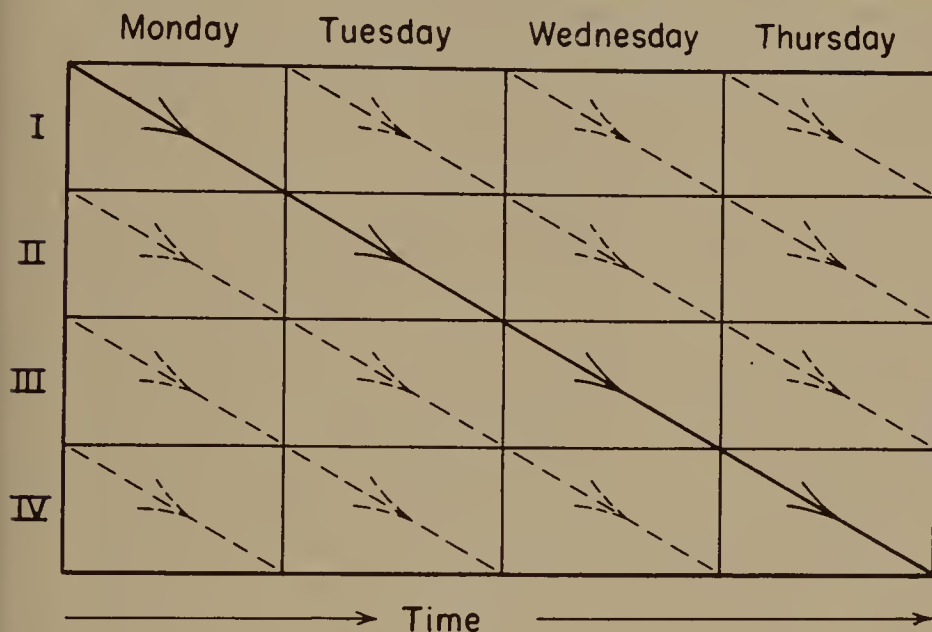


FIGURE I

of its parts. Also we want to persuade people to consume the products, which the machine will turn out, soon enough to offset deflation.

The machine analogy is implicit in Chapter 18 of the *General Theory*, and still more so in Chapter 24. It deeply colors, in my opinion, most of our "development of backward areas" thinking and it may be found in many other places. In nearly the whole "model" literature one finds, indeed, the concept of a given increase in desired output automatically and reflexively influencing the size of input and the size of equipment. "Labour working with capital upon materials to produce output." See, for example, J. R. Hicks: *The Social Framework*.

Certainly the idea of an integrated machine and of the *processing* of output is a great gain in realism and applicability. But the practical consideration of economic forecasting and economic policy has

led me to feel even the "machine" picture increasingly inadequate. Suppose we assume such an economic machine, completely built up and in a stationary, equilibrium, state. Economic production under those circumstances would be a smooth, evenly-repeated flow. Suppose there are four "stages" in production (see Figure I). We can then draw the economy as follows:

Each "stage" is exactly dovetailed into the preceding one. Whether all four stages are under the same control makes no difference. The control units in the upper stages are slavishly producing for those in the lower stages. Today's producer of raw material, to be sure, is in fact producing for a final act of consumption, four "days" away. But this makes no difference. For the upper stage producer knows that today's final sale and the one four days away are going to be identical. "Time marches on."

But now, alas, as Roy Harrod points out, the devil in the form of *net* saving steps in. The machine begins to grow in size. Immediately comes the need of "fitting" the additional capacity into the already functioning "machine." Furthermore, we will find that even if population has started to grow, and we are therefore, for the most part, merely expanding plant and output, a certain amount of redesigning is necessary for efficient operation. The machine and its operation have to be seen in much more complicated terms. And if we bring in new invention, complication increases still more. Thus the "machine" analogy must be given its most complicated and realistic formulation as follows:

"Imagine a sausage machine in full blast with people hard at work cranking out sausages. Yet imagine that at the very same time at which some men are trying to grind out sausages, other people are taking parts of the machine to pieces — adding to them, repairing them, changing them altogether . . . People trying to turn the crank will find themselves interrupted by others who want to take out the old crank and put in a new one. Or, just as things are going smoothly, someone will try to redesign the whole machine. Or some part will break down in the middle of the whole process and have to be replaced. Yet this picture of a machine, being operated, expanded, repaired, and improved, *all* at once, is a true and faithful picture of any economic system trying to grow."²

II

So far so good! But just at this point in my thinking I have found myself being switched on to quite another track. The impulse for the switch comes not from abstract reflection but from problems of practical forecasting and practical price analysis. Furthermore,

2. David McCord Wright, *A Key to Modern Economics*, (New York: Macmillan, 1955), p. 46.

the consequences of the change in approach become quite considerable.

Most people in weighing the picture of the economic system just given, will be struck most by the discontinuities and interruptions which are implied in it. But it is not discontinuities and interruptions which will be found, upon reflection, to be the really basic change. What has most basically happened is that the *demand for the product of the upper stages of production is no longer uniquely linked to the output or even the expected output of the lower stages.*

Here we may run over very briefly the reasons why the linking is no longer mechanical, and why, in a growing, changing, economy, current investment no longer need be determined *either* by current or expected consumption. The writer has elsewhere used the example of beer sales and breweries. Under what circumstances might a new brewery be built even when beer prices or beer sales are falling? There are three of them — the “better” beer, the “cheaper” beer, and the “bull-headed” brewer. The first two cases are obvious. The third case is one in which a business man believes himself to be wiser than the market. He thinks things will improve, so he builds now. And it is undeniable that in building now he *may* make them improve. Finally, and expanding the analogy, there are projects built to meet a long-run trend, e.g., roads, bridges. To these a drop in immediate consumption is a matter of indifference — even encouragement.

The full impact of what we have just said is realized only when we grasp that all these cases — the better product, the cheaper product, the “bull-headed businessman,” and the long-range type product can occur at any stage of the flow of goods through the process of production (incidentally we must remember that this flow is not a simple linear one, but filled with “loops and whirlpools”). *Anywhere*, in Figure I, for example, in the progress of production from stage I to stage IV, sudden “autonomous” increases or decreases of demand can occur. The businessman, then, who thinks only in terms of final sales to consumers, or the forecaster preoccupied with consumption, may alike find themselves caught in deep and disastrous errors.

The beauty of the machine analogy, however, even with some technical change admitted, was that it rendered the optimum output of each stage an “objective” fact determined by the *current* requirements of the next stage down. But once the decision-maker realizes the possibility of autonomous unexpected change in the lower stages of the economy, he will no longer (if he is intelligent) respond blindly to increases or decreases in current orders from below. He may hold off, for example, and wait to see if the new demand will keep up.

We will see later on that no "monopoly" power is implied in such a phenomenon. But it is just at this point that, abandoning the mechanical analogy, I have passed through into a quite different concept of the economic system.

III

What "is" the economic system? In what terms may we most accurately and usefully think of it? The answer to which practical experience and theoretical reflection have now alike brought me is that the economic system, in a growing, changing society, may be most effectively and accurately understood *not* in terms of an "equilibrium" of "spatial" forces, *nor* yet in terms of the construction of, and flow of materials through a machine, but rather as a series of what, following Whitehead, I will call "*events*." Putting the funda-

" Wednesday "



FIGURE II

mental idea more concretely, I submit that the best method of approaching the economic system is to think of it in terms of a consecutive series of roughly simultaneous *decisions* being made by a number of *coexisting* but only partially *codeterminant* decision *units*, within (conceptually) discreet intervals of *experience-and-decision-time*! The basic units for the grouping and analysis of our knowledge of economic behavior become, in other words, on the one hand the individual *decision* by the decision *unit*, and on the other the *interval* of experience-plus-decision *time* "within" which the coeval but only partly codeterminant decisions are grouped.

Perhaps it will help at this point, if we show diagrammatically the shift in emphasis which is involved in our switch from mechanical process to consecutive "event." In Figure I we showed a smooth, continuous, and repeated flow of materials through four stages of production. The important element was the study of the *flow*. But now in Figure II, instead of studying the entire flow, we isolate a

single “day” of it. Instead of showing the decision units I, II, III and IV as parts of a diagonal flow, we show them as “spatially” superimposed but “temporally” in isolation. We have ceased, in other words, to think of the path of a good from raw material to final sale and have concentrated, instead, upon a time-band drawn across the flow of goods in process.

What is the reason for this shift from flow to time interval? The answer is a simple one. In a growing, changing society the movement of materials through stages of production is no longer a smooth automatic predetermined process. Instead it is a movement which, from decision period to decision period, is “held up,” conceptually (and sometimes practically) for *reconsideration*, *replanning* and possibly *redirection* in the light of the experience of the past interval and the anticipations which each decision unit entertains concerning the future. The production decision of each decision-unit “today,” in other words, is an historical *individual*. It does not, we shall shortly see, flow mechanically from the past period, nor is the decision of a future period mechanically determined by that of the present one. When therefore we begin to study either *prediction* or output policy for the economy as a whole, or for the behavior and development of the individual “firm” or decision unit over time, we want to isolate and concentrate particularly upon precisely this phenomenon of decision, experience, *reconsideration*, *replanning* and *redirection*. And it is this process of re-evaluation which, in my opinion, is best analyzed in terms of Whitehead’s “event.”

“An event,” says Whitehead, “has contemporaries. This means that an event mirrors within itself the modes of its contemporaries as a display of immediate achievement. An event has a past. This means that an event mirrors within itself the modes of its predecessors, which are fused into its own content. An event has a future. This means that an event mirrors within itself such aspects as the future throws back on the present, or, in other words, as the present has determined concerning the future. An event has *anticipation*:

“The prophetic soul
Of the wide world, dreaming on things to come.” ”³

In this passage Whitehead has, I submit, given us precisely the framework which we need for an approach to fundamental economic analysis. In attempting to diagnose and predict the behavior of the economic system we can think of that system as a continual process

3. A. N. Whitehead, *Science and the Modern World*, p. 106 (*italics added*).

of repeated *open-ended* decision-making. The decisions are made by the decision-units, and the resulting policies come up against objective relationships and conditions, and also, to a considerable extent — but never completely — against each other.⁴ Then the decision-units in the light of the experience thus gained, re-examine and remake their plans for the next "period." But the fundamental point is that the redirection from period to period through "forecast and repetition" is never, empirically anyhow, a purely determinate process. Why so?

Each decision unit, about to act, will have available its record of past experience — which includes the past acts of the other relevant units. Obviously this past experience will greatly affect the impending new decision. Each unit will feel some degree of pressure from the coeval ideas and acts of its contemporaries. *But* each unit will *also* recombine these external stimuli in one final "decision-mix" with the *subjective, private* world of individual "hunches," aspirations, opinions and personality traits of the *men* who compose the given decision unit. No two individuals, and hence no two decision units composed of individuals, act in the same frame of reference or even exist, really, in the same *apprehended* world. There *is*, of course, a body of objective fact "today." But that body of objective fact is *variously* interpreted "today" through the psychic "lenses" of thousands of private worlds. To be sure, the decisions reached "today" through these psychic lenses meet externally to shape the objective facts of "tomorrow." But "tomorrow" is no more homogeneous or closed a system than today. For what happens "objectively" tomorrow is again variously refracted by private psychic lenses, and, still more important, is again reflexively re-examined in the light of independent *and shifting* private ideal values, before the decisions for "day-after-tomorrow's" policy are reached.

Two points are involved here and must be distinguished. First, the psychic lenses are different. But there is a still further source of spontaneous novelty (and resultant possible confusion). No human

4. Among these "objective" facts are the market relationships usually studied. I do not mean to deny their importance. But I do feel they should be studied not in isolation but as *part* of the decision-experience-redecision approach we are using. This is particularly important in view of the elements of spontaneity and futurity shortly to be stressed. So far as "equilibrium" theory goes even if we knew what the ultimate static equilibrium would be, it is a commonplace that the system is likely to experience a new change before the equilibrium has been reached. But what is less often realized is that the *motivation* and *analysis* of "adjustment to continual *maladjustment*" may be decidedly different from the simple blind maximization of present income often postulated. See below, sec. V of this paper.

decision unit, I submit, can be thought of as an entirely closed system shaped merely by its historical conjuncture and its own "personality." Nearly all units have in gréater or less degree, I believe, some ability to reach out and apprehend what I have called "novel combinations of the eternal objects" — those latent patterns of potentiality — scientific, aesthetic, moral and social — ever present in the universe but not obviously so — only latently. Man's consciousness, and here, remember, I speak purely empirically, will be found on the basis of practical experience often to have a fundamentally (in the present state of science anyhow) inexplicable ability sometimes unpredictably to transcend its surroundings and to come into rapport with latent novel patterns largely, or quite outside the known frame of its initial socio-temporal occasion. Genuine novelty enters the world when some mind registers such a contact — and such contacts can occur in many fields and in many degrees. The values among which men choose are not always those of their own time or even those of which they have heard or read!

So far as the means within our disposal are concerned, these choices and contacts are undetermined and indeterminate. What they are in the metaphysical acts of faith of the 100 per cent materialist, or in the eye of God, is not our province here. We speak here on a *practical* level and from a *practical* level, "spontaneous" novelty, inexplicable or not, is a fact that must be allowed for.⁵

IV

Back now to economics. The major limitation of the "event" approach is found in trying to select, practically, an appropriate time interval. Difficulties of selection, furthermore, bring us into contact with a group of subordinate phenomena which we must be careful not to overlook in our final picture. J. R. Commons made his basic unit not the "event" but the "transaction."⁶ What is the difference?

5. This whole problem is confused by an insidious bias in language. Whenever we "explain" language forces us to *imply* unilateral causality. Thus we say Joe Brown chose to accept such and such a value "*because*" of such and such. The implicit (and illegitimate) meaning conveyed is that Joe *had* to accept the value — that the pull of that value was so great he *could* not resist it. The fact is, however, that all we have really seen is, objectively, an *act* of choice. But the observance of the *act* of choice gives no necessary proof that a *particular* selection, among alternatives, was *inevitable* and "imposed" upon the chooser by circumstances. Yet as soon as we "explain," language forces us into precisely such an unwarranted implication. "For what he did explain he also did determinate." Language forces us to read *explanation* (i.e., determination) into what is in fact just description.

6. J. R. Commons, *Institutional Economics*.

Well, all transactions are events but not all events are transactions. The word transaction, as I see it, concentrates upon dealings *between* decision units. But our event-and-decision approach concentrates upon the formulation, *within* decision units, of policies for future action. Actions, of course, necessarily involve transactions, but not necessarily redirections of policy. Perhaps decisions can best be ranked in the degree of "futurity" and "externality" which they involve. By "futurity" we will mean the degree to which the decision is based upon expected conditions rather than existing ones. By "externality" we will mean the degree to which the decision and ensuing act is merely a passive response to external pressure rather than a free choice of an individually determined policy.

In old-fashioned classical equilibrium market economics, there is almost no conscious "futurity" or positive policy. Supplies are dumped on the market, "blindly" exchanged, and a new price "automatically" determined (as set by the marginal pairs). Only at the end of the "day" will there be any re-examination. The fishermen, say, can *then* decide whether or not to go back to sea for the *next* market-day's catch. Even here, however, there is rather less externality and blind automatism than often supposed. The bargaining ability of "good traders" may give results notably different from what the "objective" marginal pairs — if we knew them — would initially have dictated. However that may be, when we turn, on the other hand, to an act of investment — say the construction of a new plant — it will be found that that may be almost all futurity and independence, and may involve only a minimum of objective, present conceptual market competition with other factories presently producing the same product. Thus a full theory of economic activity will contain *both* the dealings *between* units, *within* a time interval, and the replanning, or deciding of production plans, *by* units during the same interval. Such replanning will be the result of the dealings between units (if there have been any) during the interval, plus other observed data, plus the individual reflexively-formed anticipations and values of the units themselves.

It follows that we cannot make our time interval so short that no experience has been gained. For in that case there would be no reason, ordinarily, to change the initial decisions. Nor do we like to make the interval so long that several stages of re-examination are lumped into one. Again, obviously, the decision intervals differ from market to market and industry to industry. Yet important as these facts are, practically speaking, I hope that we can leave them to one side for the moment in order to get a basic over-all picture.

What, from our analysis thus far, is a gross national product report for a given year with all its elaborate components and subdivisions? First of all it has *something* of the character of that "bar," shown in Figure II, drawn across the stages of production within a given period of time. However, since the GNP figures are conceptually only net, a tremendous number of decision units, stages and transactions, within the year, and of the greatest importance, are omitted.

But secondly, and leaving aside the problem of the omitted transactions and decision units, the GNP figure for a past year, may be viewed as the net *objective* residue of *all* the decisions reached within an arbitrarily bounded "event"-interval or decision-interval — the year, let us say, 1956. These output (or reported payment) statistics for a given year stand to the real *process* which they reflect in much the same relation that an empty conch shell does to the animal that built it and has now left it. The basic fallacy of much mathematico-statistical prediction is that it tries to predict the behavior of tomorrow's conches merely from the measurement of yesterday's abandoned shells!

Let us not go too far! A very great deal *can* be predicted about the behavior of future conches from the measurement of past shells. Yet even for conches there are both external accidents and internal mutations. And when we encounter humans, and decision units composed of humans — say committees — we also encounter precisely the sort of reflexive self-criticism (which may include the creative apprehension of a novel value pattern) leading to the formulation of *new* policy, which we have been describing in the past section. So the GNP report for next year, which we are trying to predict, will never be *only* the sum of past history and presently known relationships; but, on the most practical level, will also manifest the effects of spontaneous conceptual novelty and *self*-criticism on the part of some of the various decision units composing the annual "event." Since this part of the annual "event" can never be accurately foreknown we find here that element of "hunch" which inseparably clings to economic prediction.

In the machine analogy of Figure I each decision unit is a mere mechanical part, dovetailed into the whole. But the superimposed "stages" of Figure II may no longer be thought of as the tightly fitted parts of a machine — they are much more like a group of ships sailing *roughly* in the same direction. But it is not even a military convoy — all in order. Rather, one finds instead a disorderly huddle. Some steer to port for a while, while others go to starboard. They

cross each other's bows. They signal back and forth. They even sometimes ram and sink one another. Arbitrarily, we "freeze" these components (or their net effects) for a given time interval in a GNP figure. But, unhalted by our effort, the disorderly group continues to zigzag and stagger constantly onward across the temporal economic ocean — independent, self-critical, and but imperfectly communicative. By various social conventions the "masters" of the component "ships" tend to reassess their courses at somewhat similar conventional intervals. They constantly observe and feel each other's behavior, they also read each other's reports at somewhat conventional intervals and allow for what they read in what they themselves plan. We economists, in turn, try to set the time-intervals of *our* over-all reports — the GNP figures — in some relationship to those conventional decision intervals, which form the hazy and overlapping (or just arbitrary) boundaries of the reported economic "event." But the essential picture remains one of independence, spontaneity, and only the loosest of order. Such a picture — one of loose, spontaneous, repetitive decision, *such* is the economic system!

If, indeed, one wishes to cling to the machine analogy, the sausage machine must become not merely one simultaneously "being operated," being changed, etc., etc., all at once. It must further be thought of as a machine whose component parts have legs and independent wills, and which parts therefore are constantly deciding to walk around, change *themselves* and their functions, or stop altogether. The analogy has thus travelled all the way from the input-output flows of a turbine to something of the nature of a disturbed beehive or anthill.

V

It is time now as a sort of footnote to sketch a few of the implications of our picture for various points of economic theory. First of all, its relationship to the cash balance theory of velocity and of money is of the closest. Given the existence of cash balances, and balances borrowable from the banks, a shift in expectations on the part of the decision units may result in tremendous shifts in velocity and expenditure within a given period. We do not have to wait for money to "circulate." We have J. R. Common's "forecast and repetition" system.

Again in the theory of the multiplier, the effects may be either way. A quick change in anticipation and psychology produced by more buying *may* be communicated from "vessel" to "vessel" almost

instantly *within* one period. The increase in production does not *have* to wait to feel the slow turnover of period analysis. Contrariwise, however, the initial impulse of more buying may not be communicated at all. Extra purchasing power may just be hoarded if the outlook of the decision units is pessimistic.

Concerning the theory of the firm, the repercussions are still more profound. The naive picture, often entertained, is of each decision unit blindly following the orders given from the stage "below" it, and, either looking no further ahead, or else blindly assuming that what is true today will be true indefinitely. Such a picture, however, need not be valid even under pure competition. For while the purely competitive producer knows that (or behaves as if) *acting alone* he could not change the market price, this does not mean that he must be so stupid as to believe that market price will *not change*. Not at all. In "today's" production decision he decides on present production in the light of *expected* prices, *expected* technique, *expected* costs. And, be the competition never so pure, all these can change — or be expected to change. Furthermore, once we admit the concept of "good will" (which is the first degree of departure from absolute purity of competition) the decision unit no longer need maximize income in any given "event" or decision interval. For what both the pure and the impurely competitive producers are usually trying to maximize is an income *stream* — really, indeed, often an *accumulation* stream — and once good will be admitted, present sacrifices of short-run advantages may be desirable because of their effect upon the long-run income stream. How long a run depends largely upon the individual character of the decision unit. Returning to our "ships" we thus see that "today's" decision is reached more in the light of a "course" plotted over some considerable time rather than in the light of just today's conditions. Recognition of this element of *perspective* rather than immediate response adds yet more to the spontaneity, independence, indeterminacy of the whole.

Finally it may be asked to what extent does our picture apply to a growing, changing socialism. At first, it might seem not so much. Of course, the planning board will definitely have sessions of replanning and reflexive self-criticism. Witness the numerous successive Russian "Five Year" Plans. But theoretically there would only be the one such group to hold these sessions. The group of "ships" becomes a naval convoy and replanning is confined to the Admiral's cabin. But in fact this is too simple a picture. Too many emergencies come up, too many sudden demands are made, for the cap-

tains of the ship to be deprived of all discretion. There may well often be nearly as much disorder in the socialist convoy as in the capitalist one.

VI

In this article I have tried just to sketch, in outline, the elements of an over-all general picture of the economy. The picture is not complete and not intended to be. But I hope my "poetry" will be found to be of real practical usefulness in organizing our ideas.

DAVID McCORD WRIGHT.

MCGILL UNIVERSITY

THE CLASSICAL ECONOMISTS AND THE FACTORY ACTS — A RE-EXAMINATION

By MARK BLAUG

I. Introduction: The need for a reconsideration, 211. — II. The first phase of the debate, 1819–1844, 212. — III. The second phase: the Ten Hours Bill, 217. — IV. Mill's disciples and Jevons, 221. — V. Evaluation of the classical analysis, 223.

I

British historians of the Industrial Revolution are unanimously of the opinion that early factory reform was achieved in the face of strong hostility from the economic experts of the day. It does not matter whom we consult: Toynbee, Trevelyan, the Hammonds, Cunningham, Clapham; the classical economists are always depicted as unalterably opposed to the Factory Acts.¹ But if we turn to the historians of economic thought a very different interpretation emerges. Marshall, for instance, asserts that the classical economists "supported the factory acts, in spite of the strenuous opposition of some politicians and employers who claimed to speak in their name"; he cites McCulloch and Tooke as examples in point.² In his *History of Economic Analysis*, Schumpeter flatly declares that "Most 'classic' economists supported factory legislation, McCulloch especially."³ K. O. Walker, in an article which examines the question in some detail, concludes that "the direct influence of the political economists on labor legislation was negligible" and that "any influence that was exerted tended to favor, rather than oppose, the passage of the Factory Acts."⁴ Lionel Robbins' study of *The Theory of Economic Policy in English Classical Political Economy* deals briefly with this issue; he suggests that the classical authors generally favored regulation of child labor while disapproving of legislation for adults.⁵

1. For a recent example see R. G. Cowherd, *The Humanitarians and the Ten Hour Movement in England* (Boston, Mass., Kress Library of Business and Economics, 1956), pp. 5–6, 9–10.

2. *Principles*, pp. 47, 763n. See also Marshall's *Industry and Trade*, pp. 763–65.

3. (New York, 1954), p. 402.

4. K. O. Walker, "The Classical Economists and the Factory Acts," *Journal of Economic History*, I (Nov. 1941), 170. See also L. R. Sorenson, "Some Classical Economists, Laissez Faire, and the Factory Acts," *ibid.*, XII (Summer, 1952), which reaches similar conclusions. Sorenson documents the assertion made above about the opinions of economic historians.

5. (London, 1952), pp. 101–3.

How can we account for such widely divergent interpretations of what is, after all, a matter of record? One answer is that the evidence which has so far been considered is highly selective.⁶ Moreover, little attention has been paid to the successive phases of the factory reform controversy: generalizations have been advanced on the basis of writings published at different times and under distinctly different circumstances.

The fact of the matter is that the attitude of the classical writers was conditioned at each stage of the debate, by the degree of regulation that had already been achieved. Many a factory bill, whose introduction had been bitterly opposed, met with approval once it became law. Although the classical economists supported the *principle* of granting protection to children, they were aware that the unavoidable consequence was a shorter working day for adult operatives; rather than to countenance that they preferred to dispense with the benefits of regulated child labor. Thus, we are faced on the one hand with differences of opinion among the classical economists as to the desirability of further restrictions on the employment of children, and on the other hand with a general tendency towards rear-guard action designed to prevent the effective regulation of adult labor.

For this reason the question whether the classical economists did or did not favor the Factory Acts cannot be answered. This much, however, is a matter of pure academic interest. The real significance of the discussion lies in the opportunity which it affords to study the quality of classical policy-pronouncements. Were their opinions based upon economic considerations, such as the effects of shorter hours on employment and real wages, or solely upon fundamental value judgments embodied in the tenets of *laissez faire*? In the concluding section of this paper I shall attempt to evaluate the merits of the classical economists' position in the light of their own analytical apparatus and the relevant factual knowledge available to them.⁷

II

The history of factory legislation in England begins with Peel's Bill of 1819. An earlier Act of 1802, regulating the labor of parish apprentices, was an extension of the Poor Laws, not a factory act; no new power of the state was at issue. Peel's Bill, however, did raise the question of state interference in private industry; it reduced the

6. This is particularly true of Walker's analysis which deals only with the literature up to 1833.

7. The presentation of the argument has gained in clarity through the criticisms of Mr. M. Leiserson.

working day of children under sixteen to twelve hours and prohibited altogether the employment of children under nine years of age. The act applied only to the cotton factories and inadequate inspection provisions made it largely inoperative. Nevertheless, there was opposition to the bill, particularly from the House of Lords, in the form of an appeal to "that great principle of Political Economy, that labour ought to be left free." The proponents of Peel's Bill, on the other hand, defended the measure on the grounds that children were not "free agents."⁸ Economists took little interest in the debate; Malthus alone gave public support to the measure.⁹

Additional restrictions on child labor in the cotton factories in 1825 and 1831 improved but little upon the Act of 1819. But with the publication of Oastler's letters on "Yorkshire Slavery" and the appearance of Sadler's Committee Report (1833) the movement for factory reform began to assume a more radical tone. Lord Ashley's motion of a Ten Hours Bill, applicable to all persons under the age of sixteen, led to the appointment of a Royal Commission to collect further evidence. The Commissioners — Thomas Tooke, Edwin Chadwick, and Southwood Smith — proposed several amendments to Ashley's Bill to prevent interference with the free employment of adults.¹ The final version, known as Althorp's Act, limited the working day of persons between thirteen and eighteen to twelve hours a day and of those between nine and thirteen to nine hours a day.²

After the passage of Althorp's Act it was necessary to employ children in part-time relays since the work of the adult spinners and weavers depended, for technical reasons, upon the labor of their young assistants. The factory inspectors devised a variety of schemes for co-ordinating the work day of different categories of labor but none of the plans proved completely successful. The relay system soon became one of the major devices for evading legislative control. The leaders of the Ten Hours party were quick to point out that it

8. W. Smart, *Economic Annals of the Nineteenth Century* (London, 1910), I, 688, 702-3; Walker, *op. cit.*, p. 175.

9. T. R. Malthus, *Essay on Population* (5th ed.; London, 1817), p. 282. Only some dozen tracts appeared on Peel's Bill in contrast to the flood of pamphlets that accompanied the legislation of the '30's and '40's. See J. B. Williams, *A Guide to the Printed Materials for English Social and Economic History, 1750-1850* (New York, 1926), II, 192-94.

1. See M. W. Thomas, *The Early Factory Legislation* (London, 1948), pp. 55-56.

2. In addition, night work was abolished for those under eighteen, and the scope of regulation was extended to all textile factories, with the exception of lace and silk mills. Furthermore, employment for children was made conditional upon attendance at school for two hours a day and machinery of inspection was provided to supervise the enforcement of the Act.

was impossible to separate the adult from the child for purposes of legislative control; in short, they did not attempt to disguise their ulterior aim of limiting the hours of adult labor by means of placing restrictions upon the hours of children. Classical political economy, however, sanctioned a limit on the employment of children below "the age of consent" so long as this could be achieved without encroaching upon the working hours of adults. Consequently, economists arraigned themselves against the Ten Hours movement as its ultimate purpose became increasingly evident.³

The years between the Acts of 1833 and 1844 mark the first phase of the debate; at this point there was still great variety in the attitudes of individual economists. It was only in the 1840's that something like a uniform position began to emerge. Nevertheless, all the leading arguments in the controversy make their appearance at this stage of the discussion.

The first to commit himself, even before the passage of Althorp's Act, was John Stuart Mill. Writing in a popular weekly in 1832, he expressed a desire to see "a law established *interdicting* altogether the employment of children under fourteen, and *females of any age*, in manufactories".⁴ He anticipated objections to such a law drawn from the "non-interference philosophy" and admitted that he, too, was a partisan of this principle "up to a certain point." He drew attention, however, to a significant exception:

"The case in which it would be to the advantage of everybody, if everybody were to act in a certain manner, but in which it is not in the interest of any individual to adopt the rule for the guidance of his own conduct, unless he has some assurance that others will do so too. There are a thousand such cases; and when they arise, who is to afford the security that is wanted, except the legislature?"

The case of child and female labor is a typical example, he went on to say; here private and public benefit must diverge unless a universal

3. As one of the advocates of shorter hours put it bitterly:

"They could not refuse to protect children, but they are 'political economists'; and though, as men, they could no longer screw up their minds and hearts so far as to sacrifice any more limbs and lives of infants, the science would not suffer them to invade the 'freedom of industry' by involving the adults in that protection which they were obliged to give the child. It is this absurd attempt to separate the adult from the child in its labour, that has rendered every Act that has ever been passed to give protection to children almost void." C. Wing, *Evils of the Factory System Exposed* (London, 1836), p. 17 quoted by Thomas, *op. cit.*, p. 89.

4. "Employment of Children in Manufactories," *The Examiner*, Jan. 29, 1832, p. 67. The article appeared anonymously; for evidence of Mill's authorship, see *Bibliography of the Published Writings of J. S. Mill*, ed. N. MacMinn, *et al.* (Evanston, Illinois, 1945).

compact can be secured. This argument could have been applied with the same force to the labor of adult males but Mill failed to carry it through.⁵

Robert Torrens supported Ashley's Ten Hours Bill when it came up for debate in Parliament, but with an important qualification. Since the Corn Laws had raised the cost of food and thus depressed real wages, the working class was entitled to shorter hours without a reduction in money wages. Still, the tariff on agricultural produce should be lowered so as to "create a margin on which your short time might safely stand."⁶ In a work published shortly before the passage of Althorp's Act, he declared:

The evidence presented by the Royal Commission of 1832 makes it imperative on Parliament to interpose, to shorten the hours of labour, and to save the infant labourer from the cruel oppression of excessive toil. But while humanity cries aloud for such intervention, and while it must be promptly and freely granted, the truth should at the same time be declared, that a Bill for regulating the hours of labour, though framed by a consummate wisdom, cannot reach the root of the disease.⁷

The "root of the disease," of course, is the Corn Laws.

George Poulett Scrope took a similar view in his *Principles of Political Economy*: the Factory Bill is "a measure which in a healthy state of society would be a needless interference, though in the existing circumstances of the country, it seems to us highly desirable."⁸ *The Westminster Review* (under the proprietorship of Colonel Perronet Thompson, an ardent Benthamite and free trader), varied the argument: it condemned Althorp's Act as a "restrictive blunder" and depicted the Ten Hours Movement as "the stalking-horse to cover and protect — the Corn Laws and West Indian Slavery."⁹ Within a decade this became the standard reply of the Anti-Corn Law League to the factory reform movement.¹ The Corn Laws were made the scapegoat of distress in the factory districts, and cheap bread was hailed as the nostrum to remedy all ills.

While the Factory Act of 1833 was still under discussion, Lord Ashley solicited McCulloch's views on the question. McCulloch had

5. In his *Principles* (1848), however, Mill pursued the argument to its logical conclusion. See *infra*.

6. *Hansard's Parliamentary Debates*, 3d series, XV, 414-15. See also Sorenson, *op. cit.*, pp. 253-54.

7. *Letters on Commercial Policy* (London, 1833), p. 73.

8. (London, 1833), p. 51; also pp. 241, 358.

9. *Westminster Review*, April, 1833, pp. 380-81. See also G. L. Nesbitt, *Benthamite Reviewing* (New York, 1934), pp. 147-48.

1. See A. E. Bland, *et. al.*, *English Economic History: Select Documents* (London, 1919), pp. 611-12; J. Morley, *The Life of Richard Cobden* (London, 1910), pp. 166-70.

spoken approvingly of factory legislation in 1827, adding the warning, however, that "no farther interference ought, in any account, to be either attempted or tolerated."² Now he wrote to Ashley: "I would not interfere between adults and masters; but it is absurd to contend that children have the power to judge for themselves on such matters."³ McCulloch's modern reputation as a friend of factory reform is largely based upon this private communication, penned under the stimulus of the shocking disclosures of Sadler's Committee Report. It is to be noted, however, that the argument goes no further than the admission that children are not "free agents," a notion that was rapidly becoming a commonplace.

Indeed, in the pages of the *Edinburgh Review* McCulloch continued to deprecate the case for legislative control. In 1835 he devoted a major article to Ure's *Philosophy of Manufactures*, a crass apology for the factory system.⁴ "That abuses have existed in some factories is certain," McCulloch admitted, "but these have been rare instances; and, speaking generally, factory work-people, including non-adults, are as healthy and contented as any class of the community obliged to earn their bread in the sweat of their brow." He saw no reason to object to the exclusion of children under thirteen years of age from factory employment provided that they were properly looked after at home. But in view of parental attitudes among the lower classes, it was likely, he argued, that children turned out of factories would become delinquent paupers. The factory system, he observed, did imbue children with disciplined habits and allowed them to extend material assistance to their parents. Nevertheless, "the Legislature did right in prohibiting altogether the employment of children in mills under nine years of age." Lest these words give comfort to factory reformers, McCulloch hastened to add that the limitation of hours was "a matter of great nicety and difficulty"; on the whole, he concluded, the less the textile trade is "tampered with" the better.⁵ Senior's *Letters on the Factory Act* (1837) is too well known to require discussion. Its importance lies in the fact that it carried the debate out of the realm of such general considerations as the proper "age of consent," the character of parental supervision, or

2. *Edinburgh Review*, June, 1827, p. 35.

3. Quoted by Robbins, *op. cit.*, pp. 101-2. See also G. Ramsay, *An Essay on the Distribution of Wealth* (Edinburgh, 1836), pp. 102-3, for the same argument.

4. On the basis of personal experience, Ure testified that child labor in factories "seemed to resemble a sport": children, working twelve hours a day, spent nine hours in idle contemplation and "sometimes dedicated these intervals to the perusal of books."

5. *Edinburgh Review*, July, 1835, pp. 464-67.

the priority of free trade over factory legislation. Senior accepted Althorp's Act as it stood but argued that, given the cost structure of the typical textile mill, further reductions in hours would wipe out the margin of profit.⁶ Senior's thesis proved to be a serviceable argument against the extension of regulation and in the next round of discussions which took place in 1844 several members of Parliament succumbed to its logic.⁷ Senior's fellow economists, however, did not take it very seriously: *Letters on the Factory Acts* is hardly mentioned, much less analyzed, in the economic literature of the day. The records of the Political Economy Club clearly suggest that Senior's argument was not accepted by his colleagues: They objected to his unrealistic estimate of capital investment upon which his conclusions were grounded.⁸ But one of Senior's basic assumptions, that output would fall proportionately with the reduction of hours, was not challenged and became an essential feature of the classical analysis of factory legislation.

III

A new Factory Act was passed in 1844 which lowered the working hours of children to six and one-half hours and that of "young persons" (boys below eighteen and girls below twenty-one) to twelve hours. This Act proved to be a steppingstone to the Ten Hours Bill of 1847 which finally secured a fifty-eight hours' week for "young persons" and for women of all ages. The passage of both measures was accompanied by an intense discussion that marked the high point of three decades of debate. Economic arguments became more concrete and were now clearly divorced from the precept of noninterference. However, there were no dramatic conversions to the Ten Hours camp.⁹ At the Political Economy Club in 1844, Edwin Chadwick put up this question for debate: "Is legislative interference

6. Contrary to popular belief, fostered by Marx's attack, Senior did not advance a general theory that profits are produced in the "last hour." Even on his own assumptions, Senior's calculations actually show no more than that a shortening of the working day by one hour would cause profits to fall from 10 to 8 per cent, given a constant output per man-hour. See K. Wicksell, *Lectures on Political Economy* (New York, 1934), I, 194-95.

7. See A. E. Bland, *et. al.*, *op. cit.*, pp. 605-6.

8. See Walker, *op. cit.*, pp. 171-72.

9. There is some evidence that Dr. Thomas Chalmers, a leading Scottish divine and author of several economic treatises, was finally won over by the Ten Hours campaign in 1847. If so, Chalmers was a singular exception. See C. Driver, *Tory Radical. The Life of Richard Oastler* (New York, 1946), pp. 476-79.

between the Master and the Adult labourer, to regulate the hours of work, expedient?" The diary of one of the participants reveals that Charles Buller, the radical philosopher, was the only member to vote in favor of such interference.¹ Chadwick, Senior, Torrens, and Tooke answered the question in the negative. McCulloch admitted much of Buller's reasoning but thought the matter could not be settled in general terms. The views of John Stuart Mill at this point are not clear; but in an article on "The Claims of Labour" for the *Edinburgh Review* (1845) he referred to the Ten Hours Bill as falling into the category of "quack schemes of reform."

The prevailing economic argument against the Ten Hours Bill is set forth in Torrens' *Letter to Lord Ashley* (1844), a curiously neglected work.² Torrens begins his discussion with a strong condemnation of the principle of "leaving things to their course." The concept of "free agents," however, is not mentioned at all. His analysis is largely concerned with "the delusion" of the operatives that "upon the passing of a Ten Hour Bill, they would receive the wages of twelve hours for the work of ten." Torrens lays it down as an incontrovertible fact that "the rate of profit in this country is already approaching the minimum at which no margin remains for an advance of wages"; "capital to an enormous amount already emigrates from our shores."

Torrens' conclusion is that the Ten Hours Bill would check production and diminish wages: "Enact your Ten Hours Bill and one of two events must inevitably ensue: — the manufactures of England will be transferred to foreign lands, or else the operatives must submit to a reduction of wages to the extent of 25 per cent."³

There is no mention in Torrens' pamphlet of the possible productivity effects of a shorter working day. Yet this had long been a favorite argument of the factory reformers. Robert Owen had testified in 1818 before Peel's Committee that a reduction from fourteen to twelve hours a day in his factory at New Lanark had actually

1. *Proceedings of the Political Economy Club, 1821-1920* (London, 1921), VI, 287-88.

2. Sorenson (*op. cit.*) contends that Torrens was definitely sympathetic to factory legislation. The evidence for this comes from Torrens' Parliamentary speeches in the 1830's while *Letter to Lord Ashley*, the most important of Torrens' writings on the Factory Acts, is not considered.

3. *A Letter to Lord Ashley* (London, 1844), pp. 64-65, 71-73. Torrens' argument was reproduced in the popular journals: see the article on "Protection of Labour," *The Economist*, April 6, 1844. Typically, however, *The Economist* based its case on laissez faire (see S. Gordon, "The London *Economist* and the High Tide of Laissez Faire," *Journal of Political Economy*, LXIII (Dec. 1955), 478, 483).

resulted in an increase of output.⁴ Speaking in the House of Commons in 1844, Lord Ashley recalled Owen's testimony by way of an attack on Senior's "last hour" theory. Reviewing the successive Factory Acts since 1819, he pointed out: "you had no diminution of produce, no fall in wages, no rise in prices, no closing of markets, no irresistible rivalry from foreign competition, although you reduced your hours of working from 16, 14, 13, to 12 hours a day."⁵ The implication is that productivity per man-hour had risen with each reduction in the length of the working day.

Ashley's argument is loose, of course: dynamic factors unrelated to shorter hours might account for the facts. The same argument, however, more carefully stated, appears in a popular treatise of the forties, William Thornton's *Over-Population and Its Remedy*. Thornton reviewed the whole question in the light of the imminent repeal of protection. If "the daily labour of British operatives were shortened," he thought it "very possible that their wages would fall." But once the Corn Laws were abolished, lower food prices might leave real wages constant, or even raise them, despite the fall in money wages owing to a Ten Hours Bill. Moreover,

It is not quite certain that a diminution of produce would result from shortening the duration of labour. Persons who are not obliged to work so long may work harder than before, and may get through the same quantity of work in a short time as formerly occupied them for a longer period. . . . If so, the limitation of labour to ten hours daily would not in any circumstances reduce wages, and at all events the reduction might be either prevented or neutralised by the establishment of free trade in food.⁶

Unhappily, Thornton's analysis made no impression on his contemporaries. McCulloch, for example, continued to discuss the regulation of hours along traditional lines. "We should be inclined to think," he wrote in 1846, "that the existing regulations respecting factory labour in this country are about as reasonable and judicious as they can be made." Then he went on to praise Torrens' *Letter to Ashley* as "the best tract in opposition to the ten-hours project."⁷ In the fourth edition of his *Principles* (1849) he added a few pages on the Factory Acts, lauding the Act of 1844 as consistent with "claims of humanity" and "the interest of manufacturers" but

4. See B. L. Hutchins and A. Harrison, *A History of Factory Legislation* (London, 1911), pp. 19-23.

5. *The Ten Hours Factory Bill. The Speech of Lord Ashley, M.P. in the House of Commons on Friday, May 10, 1844* (London, 1844), pp. 15-16.

6. (London, 1846), p. 399.

7. *The Literature of Political Economy* (London, 1846); London Reprints No. 5 (1938), pp. 294-96.

roundly condemning the Bill of 1847 because it tended to restrict the hours of adults. At this point he turned to a new argument. The conditions of the working class, he declared, rest ultimately upon the size of the wages fund relative to population; the real issue, therefore, is not whether eight, ten, or twelve hours constitutes the "proper" length of a working day.

If . . . the longer be introduced by the customs of the country, in preference to a shorter period, it is a proof that there is, if not an excess, at all events an extremely copious supply of labour; and that the labourers are, in consequence, obliged to submit to the drudgery of lengthened service . . . it is difficult to perceive how the hours of work . . . should be lessened by a legislative enactment without at the same time, and by the same act, reducing wages.⁸

John Stuart Mill touched briefly on the economic objections against the Factory Acts in his *Principles* (1848). Whether a reduction of hours without a cut in wages would inevitably displace labor was, he said, "in every particular instance a question of fact, not of principle." For the most part his analysis of factory legislation dealt with the propriety of government intervention along the lines laid down in his earlier article of 1832.⁹ If a nine hour day were proved to be in the interest of the working class, Mill reasoned, state action would be required "not to overrule the judgment of individuals respecting their own interest, but to give effect to that judgment." He concluded: "I am not expressing any opinion in favour of such an enactment . . . but it serves to exemplify the manner in which classes of persons may need the assistance of law, to give effect to their deliberate collective opinion of their own interest." He condemned the Acts of 1844 and 1847, however, on the grounds that they excluded working-women from factories, although women were "free agents" as much as men.¹

8. *Principles of Political Economy* (London, 4th ed., 1849), pp. 427-30. See also McCulloch's *Treatise on the Circumstances Which Determine the Rate of Wages* (London, 1851), pp. 93-97, and *Treatises and Essays* (Edinburgh, 1859), pp. 453-54.

9. Mill's argument here is nothing but an early example of Pigou's famous distinction between private and social costs, as W. J. Baumol pointed out: *Welfare Economics and The Theory of the State* (Cambridge, Mass., 1952), pp. 15-16, 150-52.

1. J. S. Mill, *Principles of Political-Economy*, Ashley edition, pp. 964-65, 959. Senior took the same view on female labor: *Industrial Efficiency and Social Economy*, ed. S. L. Levy (London, 1929), II, 307-8.

IV

The Ten Hours Bill of 1847 had failed to abolish the system of employing children in part-time shifts; consequently, it was possible to keep adult male operatives at the bench for fifteen hours a day without violating the letter of the Act of 1847. Renewed agitation at last secured the "normal working day" for women and children in 1853: hours of legal employment and meal times were specified in greater detail so that it became difficult to employ relays. The scope of the Ten Hours Bill was extended in the 1860's, although industries other than textile were not covered until the Consolidating Act of 1878. The minimum age of child labor was now raised to ten, the employment of women was further restricted, and sanitary inspection and safety-regulations were improved. None of this legislation, except the details of sanitation, was applicable to adult males but their weekly hours, of course, were almost everywhere scaled down to sixty or less.

Meanwhile, fragmentary statistical data on the effect of the Act of 1847 had been gathered by the factory inspectors. The initial consequences were partly obscured by a severe trade depression. Wages in textiles fell, but much less than the 16 per cent reduction in hours or the 10 per cent reduction in piece rates. After the revival of prosperity in the 1850's, Horner and Tooke declared that the Ten Hours Bill had not depressed either earnings or output owing to an increase in the intensity of labor.²

There is no indication that economists shared the belief that shorter hours had paid for themselves through a rise in output per man. New editions of Mill's *Principles* in the fifties and sixties reveal no alterations with respect to the topic under discussion. Cairnes' writings contain no explicit discussion of the Factory Acts. Fawcett, however, delivered a lecture on the question in 1872 in the midst of a new campaign for a nine hours' day. At the outset he expounded the familiar theme of the free agent:

It certainly appears to me that it is quite as desirable to pass a law limiting the number of hours which a child is permitted to work, as it would be undesirable to impose similar restrictions upon men and women. If grown-up persons overwork they do it of their own free will.

Moreover, he had no patience with Mill's "hypothetical argument" in favor of state intervention. This is "the old story," Fawcett com-

2. See G. H. Wood, "Factory Legislation, considered with reference to the Wages, etc., of the Operatives Protected thereby," *Journal of the Royal Statistical Society*, LXV (June 1902), p. 297.

plained, which requires us to believe in the collective wisdom and infallible judgment of the legislature. He proceeded to examine the notion that a diminution of hours could increase the efficiency of labor and thus leave output unaffected. He admitted that there was some factual evidence which might be adduced in behalf of this argument. Still, he insisted that generally entrepreneurs could be trusted to maximize profits and, thereby, to achieve an optimum length of the work day from the viewpoint of maximizing output per man-hour.³

The success of the Nine Hours Movement, Fawcett warned, would open the way to a campaign in favor of eight hours, and so forth. Already, England "can scarcely hold her own in some trades in which she once had an almost undisputed supremacy." When the Nine Hours Law came up for debate in Parliament Fawcett spoke against it on the grounds that "this House has no right to interfere with the labour of adults" or to place the employment of women on a different footing from the employment of men.⁴

Although Jevons is not a classical economist, his treatment of the Factory Acts contains some instructive differences as well as similarities to the classical analysis. Jevons denied, first of all, that the question can be decided once and for all on "some supposed principle of liberty." The same principle, if it existed, would apply to adult women whose hours were already regulated. Moreover, a mass of "paternal legislation," such as the Truck Acts, the Coal Mines Act, and a series of bills relating to merchant shipping and the fencing of machinery, had long ago been sanctioned for the protection of adult men. On the face of it, he saw no reason to prohibit state action in the matter "if it could be clearly shown that the existing customs are injurious to health and there is no other probable remedy."⁵

At the same time, Jevons' analysis is quite innocent of the type of consideration introduced by Thornton.⁶ Jevons believed it to be "an economic fallacy" to suppose that shorter hours could give rise to any counterbalancing advantage other than the workmen's enjoyment of more leisure.⁷ Then, ignoring Mill's contention that private interests were fundamentally interdependent, he concluded:

3. H. Fawcett, *Essays and Lectures on Social and Political Subjects* (London, 1872), pp. 36, 113-15, 120.

4. H. Fawcett, *Speeches on Current Political Questions* (London, 1873), pp. 122 ff.

5. W. S. Jevons, *The State in Relation to Labour* (London, 1882), p. 65.

6. Thornton's argument was finally "rediscovered" by Marshall (*Principles*, pp. 695-96).

7. See Jevons, *The Theory of Political Economy* (2d ed.; London, 1879), pp. 63-64, and *Methods of Social Reform and Other Papers* (London, 1883), p. 109.

When we observe too, that trades unions are already constantly wrangling with employers for a reduction of hours, while individual workmen are generally ready to work overtime for a moderate inducement, we shall be led to think that there is no ground whatever for legal limitation of adult male labour in the present day.

V

The classical analysis of the Factory Acts consisted of two quite separate strands of thought. On the one hand, factory legislation was criticized in terms of the doctrine of "freedom of contract" between enlightened economic agents. On the other hand, it was held that something like a Ten Hours Bill would spell the ruin of British industry if unaccompanied by a drastic fall in money wages. We will examine each argument in turn.

In so far as the problem was treated as a matter of enlightened individualism, the attitude of the classical economists was unambiguous: where self-interest was plainly unenlightened, as in the case of children, they recommended intervention by the state, differing only about the proper age of consent and the scope of parents' right of supervision. Nevertheless, in practice this meant that they acquiesced in just so much legislation as had already been achieved; at each stage of the debate they warned against further measures. Invariably, notions about the age at which a worker becomes a "free agent" changed in the wake of legislation, at each turn approving a *fait accompli*.

McCulloch's treatment of the question is typical in this respect. One would hardly describe him as a supporter of the Factory Acts. Senior is another telling example. In his *Letters on the Factory Acts* he agreed that no child of eleven should be employed as much as twelve hours a day; this implied acceptance of Althorp's Act which defined thirteen as the age at which "the period of childhood, properly so called, ceases." In 1841 he thought that the "age of consent" ought to be raised from thirteen to fourteen; in 1847 he urged that it be set at sixteen, that is, two years below the age of consent stipulated in the Ten Hours Bill. Similarly, he now assented to a six and one-half hour day for children, as called for in the Act of 1844. But he never changed his mind about the undesirability of regulating adult labor.⁸

8. Sorenson, *op. cit.*, pp. 260-61. Walker's observation (*op. cit.*) that "reputable and orthodox economists like Colonel Robert Torrens, Joseph Hume, Thomas Tooke, Edwin Chadwick, and Leonard Horner, were all favorable to factory legislation as long as it was limited to children" completely begs the question. Not only were some of these "economists" never regarded, by themselves or others, as spokesmen of economic science, but all public figures after 1820 or thereabouts approved of factory legislation limited to children below some age or other.

Apart from being wise too often after the event, the classical economists never faced the question whether it was, in fact, possible to protect women and children without interfering with the employment of adult males. Strictly speaking, economists are not concerned with administrative feasibility. Still, the total neglect of the difficulties created by the relay system rendered most of the classical prescriptions for legislation void of practical significance. In addition, the notion of "free agents" was in itself extremely vague. The whole case against the Factory Acts based on this concept falls to the ground once we consider Mill's argument that the ability of adult operatives to recognize their own self-interests does not prevent them collectively from working longer hours than each alone might have found desirable. Although Mill presented this argument in one of the most widely read treatises of the period, he never for one moment succeeded in deflecting the debate from the well-worn theory of free agents. This is all the more surprising since this doctrine is repeatedly attacked in the reports of the factory inspectors.⁹

In the case of Mill and Fawcett the problem of factory reform was complicated by the issue of feminism. They feared that the Ten Hours Bill would encourage the substitution of unprotected adult males for protected female workers.¹ Since the emancipation of women was held to be dependent upon unlimited access to factory employment, they thought it necessary to condemn the Factory Acts in so far as these involved restrictions upon the hours of women workers.

All things considered, the Ten Hours camp was not far wrong in regarding "political economy" with its slogan of "free agents" as a major obstacle to factory reform. This is even more true when we consider the arguments based directly upon economic theory. It cannot be doubted that the Ten Hours movement would have met with much less hostility if economists had insisted from the outset, as did John Stuart Mill in 1848, that the wage and employment effects of shorter hours were "in every particular instance a question of fact, not of principle." To be sure, economic theory added very little in the way of theoretical analysis to popular thinking about the Factory Acts. The level of formal analysis barely rose above the

9. See the citations by K. Marx, *Capital* (New York, 1939), p. 288.

1. Their fears seem to have been unfounded. Available data covering the period 1835-70 reveal a steady tendency to replace protected children with similarly protected adults and young persons; women above thirteen comprised from 50-56 per cent of the labor force throughout the period. See Wood, *op. cit.*, pp. 310-11.

commonplace: no effort was made to distinguish the short-run and long-run effects of a change in hours, without which distinction any analysis was bound to be naïve. In this sense, it is true to say that "had there been no classical economic theory, the arguments would have been essentially the same."² Nevertheless, the assumption of a constant productivity of labor irrespective of the length of the working day had been challenged by at least one economist, William Thornton. On the face of it, there is nothing in classical theory which would have prevented a consideration of this factor; once introduced there is little left of Torrens' *Letter to Ashley*, "the best tract in opposition to the ten-hours project."

McCulloch's use of the wages fund doctrine to show that it is fruitless to restrict hours by legislative enactment is simply wrong. He failed to realize that at bottom, and apart from humanitarian motives, the leaders of the Ten Hours movement were trying to restrict the supply of labor in order to maintain the rates of wages in periods of severe unemployment. It is no accident that all the Factory Acts in the first half of the nineteenth century were passed after vigorous working class agitation "at, or close to, a low point in cyclical fluctuations."³ At such times employers were more inclined to accept restrictive legislation, but that is not the point. Only under depressed conditions can workers hope to gain instantly by an elimination of child and female labor; in a boom the immediate effect would be a reduction in real income per family. Needless to say, the wages fund doctrine is quite adequate to show why a reduction in the labor supply does tend to reduce wages. At the same time, it must be said that the doctrine is really inappropriate to a discussion of the Factory Acts: it assumes that the size of the labor force is a constant proportion of the total population, thus ignoring variations in the child and female participation rate.

In a class by itself is Fawcett's contention that pecuniary motives alone bring about the adoption of a work day that optimizes output per man-hour. This argument is open to the objection that it assumes perfect foresight. Contrariwise, Thornton's thesis amounts to a denial of perfect knowledge on the part of the entrepreneur. We should say today that entrepreneurs may have little incentive to reduce hours since the immediate effect, if wages are kept constant, is to increase costs and decrease output; whereas, a simultaneous reduction in wages under these circumstances is bound to affect

2. Walker, *op. cit.*, p. 177.

3. W. W. Rostow, *British Economy of the Nineteenth Century*, p. 118.

efficiency adversely. Thus, employers may fail to maximize output per man-hour owing to an excessive emphasis on profit maximization in the short run.⁴ Be that as it may, Fawcett's line of reasoning clearly shows where the classical economists' treatment of hours legislation is deficient: they had no theory of the firm.⁵

There is a simple moral in all this: for some purposes a theory of economic growth is not enough.

MARK BLAUG.

YALE UNIVERSITY

4. See J. R. Hicks, *The Theory of Wages*, pp. 104-10. Even on the assumption of perfect foresight, this is a clear case of private costs diverging from social costs. There is no reason why the classical economists could not have considered this possibility; the distinction between private and social costs is implicit in Adam Smith's discussion of public works.

5. Since Jevons likewise had no theory of the firm he was unable to improve upon classical analysis in this respect.

ADAPTIVE EXPECTATIONS AND COBWEB PHENOMENA

By MARC NERLOVE*

I. Introduction, 227. — II. Åkerman's argument, 229. — III. Reformulation in terms of adaptive expectations, 230. — IV. An alternative but equivalent formulation, 233. — V. Some empirical results, 236. — VI. More complicated models, 238. — VII. Conclusions, 240.

I. INTRODUCTION

In a recent article in this *Journal*, Åkerman suggests that: "Cobweb phenomena of the traditional type involving growing disequilibrium are . . . in reality rather improbable."¹ Åkerman asserts that other writers on ". . . cobweb phenomena have operated with one unique normal supply schedule without distinguishing between short- and long-term schedules. This is the reason why they have ascribed quite exaggerated properties to the cobweb phenomena."² The purpose of this article is to examine Åkerman's proposition in the light of recent developments in the theory and estimation of distributed lags. Some tentative empirical results are presented which suggest that instability may exist, however "improbable" it may appear to be from a theoretical point of view. It is thus demonstrated once again that it is difficult to draw conclusions about reality from a purely theoretical model.

In his article on the cobweb theorem, Ezekiel set out to show that ". . . classical economic theory rests upon the assumption that price and production, if disturbed from their equilibrium, tend to gravitate back toward that normal. The cobweb theory demonstrates that, even under static conditions, this result will not necessarily follow."³ The conditions under which the cobweb theorem was supposed to apply were:

"(1) where production is completely determined by the producers' response to price, under conditions of pure competition (where the producer bases plans for future production on the assumption that present prices will continue, and that

* I am indebted for helpful comments to R. J. Foote, formerly of the U. S. Agricultural Marketing Service, and to Max Steuer.

1. Gustav Åkerman, "The Cobweb Theorem: A Reconsideration," this *Journal*, LXXI (Feb. 1957), 158.

2. *Ibid.*, p. 155.

3. Mordecai Ezekiel, "The Cobweb Theorem," this *Journal*, LII (1938), 279.

his own production plans will not affect the market); (2) where the time needed for production requires at least one full period before production can be changed, once the plans are made; and (3) where the price is set by available supply."⁴

Ezekiel does not state explicitly the period to which his supply curve is supposed to refer. Buchanan has pointed out that it must be "... completely reversible throughout its whole length with respect to each period . . ." and it therefore has at least one characteristic of a "long-period competitive supply curve."⁵ As Åkerman points out, we cannot be sure that Ezekiel had in mind any particular period, long or short.

One of the crucial conditions under which the cobweb theorem is supposed to apply — to which we return in Section III — is that "... the producer bases his plans for future production on the assumption that present prices will continue." If this is not the case, rigid price expectations may produce irreversible supply conditions in the short run.

Åkerman points out that Ezekiel defines supply during the consumption year as the aggregate of the year's harvest and the whole existing normal stock so that the very shortest-run supply schedule is perfectly inelastic. In his own model Åkerman treats *market* supply so that supply within the year is not perfectly inelastic. For our purposes, we assume that farmers face an excess demand curve for net additions to stocks and consumption, i.e., we deduct the schedule of excess supply from storage from the demand schedule for consumption in order to arrive at the derived demand for a particular crop at the farm level. Neglecting abandonment, we thus have a perfectly inelastic short-run supply schedule facing a derived demand schedule. Under these conditions and the other conditions outlined above, the following simple exposition of the cobweb theorem can be given:

Let q_t^D = the quantity demanded, q_t^S = the quantity supplied and P_t = the price, where all are taken for the period t . In the neighborhood of any point, say equilibrium, we can approximate the supply and demand schedules by linear functions:

$$(1) \quad \begin{aligned} q_t^D &= a + bP_t \\ q_t^S &= c + dP_{t-1} . \end{aligned}$$

Condition (3) above implies that $q_t^D = q_t^S$ for all t . Hence

$$(2) \quad P = \frac{c - a}{b} + \frac{d}{b} P_{t-1} .$$

4. *Ibid.*, p. 272.

5. Norman S. Buchanan, "A Reconsideration of the Cobweb Theorem," *Journal of Political Economy*, XLVII (Feb. 1939), 68-69.

The equilibrium price is one at which $P_t = P_{t-1}$. Let P_0 be the equilibrium price, then

$$(3) \quad P_0 = \frac{c - a}{b - d}.$$

If the price is initially $P \neq P_0$, will the price approach P_0 , fluctuate endlessly around it, or oscillate in explosive fluctuations? Equation (2) is a first order difference equation in actual price. Its solution is

$$(4) \quad P_t = \frac{c - a}{b - d} + \left(\frac{d}{b}\right)^t \left[P - \frac{c - a}{b - d}\right],$$

where $P \neq P_0$ is the initial price.⁶ It can easily be seen that only when $\left|\frac{d}{b}\right| < 1$ will the price tend toward equilibrium. Since $b < 0$ and $d > 0$, typically, only oscillations occur: they are explosive if $\frac{d}{b} < -1$; steady if $\frac{d}{b} = -1$; and damped if $-1 < \frac{d}{b} \leq 0$. This is essentially what the traditional cobweb theorem states.

II. ÅKERMAN'S ARGUMENT

Åkerman considers a special crop A , the demand for and supply of which have been in equilibrium for some time past. As noted earlier, he also assumes the existence of stocks, and considers the excess supply of stocks to be added to the producers' supply in arriving at the market supply. Suppose now that the demand curve for A shifts to the right. Initially, the price will rise, and, because stocks are included in supply, equilibrium of a sort will be achieved at the intersection of a less than perfectly inelastic short-run supply schedule and the new demand schedule. But, adds Åkerman, this short-run supply schedule is not the only supply schedule involved in our problem:

"A crop product A is very seldom cultivated as the sole crop of a farm. Different crops are generally combined in rotation systems. Such systems usually enhance the effectiveness of fertilizers and the possibilities of suppressing weeds, and also facilitate a more even distribution throughout the year of farm labor. A permanent enlargement of crop A , therefore, will generally be of some disadvantage and cause some additional cost. *Such cost will be undertaken only if the higher price p_a of A can be expected to be of a permanent nature* [italics mine].

6. See R. G. D. Allen, *Mathematical Economics* (New York: St. Martin's Press, 1956), chap. 6, pp. 176-208.

The extension of cultivation of A may also be attained through an extension of total land under cultivation. But such additional land will generally be of somewhat inferior quality [although not necessarily inferior, since a question of comparative rather than absolute advantage is involved] causing some additional cultivation cost and requiring a somewhat higher price p^a . The *long-term normal supply* curve for extended A production is thus of a slowly rising nature.

"In the space between the market supply curve and the long-term normal supply curve a *short-term normal supply* curve is situated. Long-term normal supply conditions will generally not be fully established until several years have elapsed after each particular price change."⁷

Thus, in the year following a shift in demand, production will expand along a "short-run normal" supply curve intermediate to the short-run and "long-run normal" supply curves.

Similarly, when price falls after production is greatly increased, production will not immediately be cut back to its former level or less than its former level, but will follow a new "short-run normal" supply curve, the elasticity of which is less than that of the "long-run normal" supply schedule but greater than that of the short-run supply schedule. Thus Åkerman's assumption of a series of "... different, successively arising and vanishing, moderately elastic short-term normal supply curves ..." leads him to a graphical demonstration that cobweb fluctuations involving growing disequilibrium are improbable.⁸

In subsequent sections of this article, the difference between short- and long-run supply or demand is expressed in analytical rather than graphical form, which makes possible more definite conclusions and leads readily to empirical applications.

III. REFORMULATION IN TERMS OF ADAPTIVE EXPECTATIONS

Åkerman pointedly remarks that a position along the "long-term normal" supply schedule will be taken only if the price corresponding to that position "... can be expected to be of a permanent nature." Later he remarks that if "... the price p_a has risen the farmer, generally, will not be convinced it will remain so elevated until several years have elapsed."⁹ Thus it might be said that farmers react to expected normal price and that this is not generally the same as price at harvest.¹

7. Åkerman, *op. cit.*, p. 154.

8. *Ibid.*, p. 158.

9. *Ibid.*, p. 154.

1. For some estimates of the elasticities of supply of certain agricultural commodities based on this principle see my "Estimates of the Elasticities of Supply of Selected Agricultural Commodities," *Journal of Farm Economics*, XXXVIII (May 1956), 496-509, and my "Estimates of the Elasticities of Supply of Corn, Cotton, and Wheat," unpublished Ph.D. dissertation, The Johns Hopkins University (Baltimore: 1956).

Expectations of future prices may be autonomous, induced, or divided into two components, one of which is autonomous, the other induced. Induced expectations are the result of movements in past prices.² Only induced price expectations are amenable to economic analysis in the present context. That a price change must remain in effect for a considerable period of time before farmers will become convinced of its permanence and take up positions along a "long-term normal" supply schedule would appear to indicate that Åkerman has in mind a particular kind of induced expectations: Farmers take past prices into account when forming their expectations of future "normal" price, but they do not give all the weight to one particular price. When current price increases, farmers may be expected to discount some of the increase, i.e., they will not believe the permanence of the entire change. Arrow and I have called such induced expectations "adaptive."³

Intuitive considerations of the sort which Åkerman sets forth would lead us to suppose that producers' supply depends on a great many past prices, i.e., the effect of a price change is distributed over a great many periods.⁴ The distribution of the effects of a price change over many periods due to expectational rigidities may be represented by a variant of a model developed originally by Cagan.⁵ Let P_t^* be expected "normal" price in period t , and let β be a constant of proportionality called the coefficient of expectations; then the model in question is

$$(5) \quad P_t^* - P_{t-1}^* = \beta [P_{t-1} - P_{t-1}^*], \quad 0 < \beta \leq 1.$$

In words, farmers revise their previous expectations of "normal"

2. Alain C. Enthoven and Kenneth J. Arrow, "A Theorem on Expectations and the Stability of Equilibrium," *Econometrica*, XXIV (July 1956), 288-93.

3. Kenneth J. Arrow and Marc Nerlove, "A Note on Expectations and Stability," Technical Report No. 41, Department of Economics, Stanford University, Stanford, California (March 25, 1957); to be published in revised form in *Econometrica* (April, 1958).

4. We thus have a distributed lag of an expectational nature. Such lags are discussed at length in my *Using Distributed Lags in the Analysis of Demand for Agricultural and Other Commodities*, forthcoming publication of the U. S. Department of Agriculture.

5. Phillip Cagan, "The Monetary Dynamics of Hyper-Inflation," in *Studies in the Quantity Theory of Money*, ed. M. Friedman (Chicago: University of Chicago Press, 1956). It should be noted that Cagan's model may be derived from Hicks's definition of the elasticity of expectations. See J. R. Hicks, *Value and Capital*, 2d ed., p. 205. In its logarithmic form the model may be derived by setting the elasticity of expectations equal to a constant (see Arrow and Nerlove, *op. cit.*). However, it is doubtful that Hicks actually had the model in mind.

price in each period in proportion to the difference between actual price and what was previously considered to be "normal."⁶

The model of expectation formation represented by equation (5) may be applied to the problem of instability in a single market.⁷ Let

$$(6) \quad q_t^D = a + bP_t$$

be the demand equation, as before. Let the supply equation, however, be

$$(7) \quad q_t^S = c + dP_t^*$$

Lagging equation (7) one period, substituting for q_{t-1}^S in (5), solving for P_t^* , and substituting the result in (7), we have

$$(8) \quad q_t^S = c\beta + d\beta P_{t-1} + (1 - \beta)q_{t-1}^S.$$

Since we assume the market always to be in short-run equilibrium

$$(9) \quad q_t^D = q_t^S = q_t, \quad q_{t-1}^D = q_{t-1}^S = q_{t-1}, \dots$$

By (9), and from equation (6) lagged one period and equation (8) we have

$$(10) \quad \begin{aligned} q_t &= c\beta + d\beta P_{t-1} + (1 - \beta)[a + bP_{t-1}] \\ &= (c - a)\beta + a + [(d - b)\beta + b]P_{t-1}. \end{aligned}$$

From (6), (9), and (10), therefore,

$$(11) \quad P_t - \left[\left(\frac{d}{b} - 1 \right) \beta + 1 \right] P_{t-1} = \frac{(c - a)\beta}{b}.$$

Equation (11) is a first order difference equation in P_t ; it may be solved in terms of the initial conditions and the constant parameters a , b , c , d , and β . If P_0 is the equilibrium price, and $P \neq P_0$ is the initial price (resulting, say, from a shift in the demand curve), then the solution of (11) is

$$(12) \quad P_t = P_0 + (P_0 - P) \left[\left(\frac{d}{b} - 1 \right) \beta + 1 \right]^t.$$

6. Equation (5) is a first order difference equation in P_t^* :

$$P_t^* - (1 - \beta)P_{t-1}^* = \beta P_{t-1}$$

Regarding P_t as a known function of time this may be solved to obtain

$$P_t^* = \beta P_{t-1} + (1 - \beta)\beta P_{t-2} + (1 - \beta)^2 \beta P_{t-3} + \dots$$

Thus, if supply depends on expected "normal" price, it depends on past prices taken with a distributed lag. The distribution of lag depends on the parameter β .

7. For an application in the case of multiple markets, see Arrow and Nerlove, *op. cit.*

The second term on the right will go to zero as t increases if, and only if,

$$(13) \quad \left| \left(\frac{d}{b} - 1 \right) \beta + 1 \right| < 1.$$

Thus, (13) is a necessary and sufficient condition that return to equilibrium will be achieved. (13) may be rewritten

$$(14) \quad 1 - \frac{2}{\beta} < \frac{d}{b} < 1,$$

since $\beta > 0$ by assumption. (14) should be contrasted with

$$(15) \quad -1 < \frac{d}{b} < 1,$$

which is the traditional cobweb result. (14) reduces to (15) when $\beta = 1$. When $\beta = 1$, equation (5) also reduces to

$$(16) \quad P_t^* = P_{t-1}$$

so that the model given by equations (5), (6), and (7) reduces to the ordinary cobweb model in the case where the coefficient of expectations, β , is unity.

The ranges of $\frac{d}{b}$ compatible with stability are shown in Figure I

for both the traditional theory of the cobweb and the model developed above. The only range compatible with stability according to the traditional theory is the cross-hatched area lying between the lines parallel to the β -axis through 1 and -1 . The range compatible with stability if β is not 1 includes the doubly cross-hatched area as well. It should be noted that the permissible range widens as β falls; almost any positive elasticity of supply and negative elasticity of demand is compatible with stability when the coefficient of expectation is low.

While it can be seen from Figure I that the possibility of stability is much improved when adaptive expectations are assumed, it is by no means certain that stability will always occur. Whether or not it will, depends on the relationship among the slopes of the demand and "long-term normal" supply curves and the coefficient of expectations.⁸

8. This relation is examined empirically for corn, cotton, and wheat in Section V.

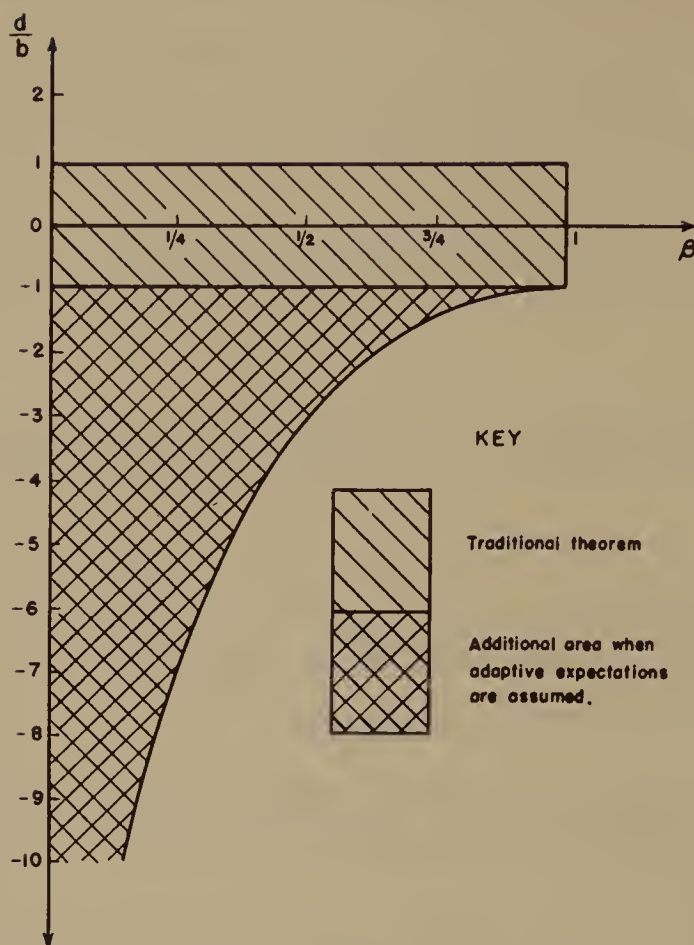


FIGURE I

RANGES OF d/b COMPATIBLE WITH STABILITY UNDER THE TRADITIONAL COBWEB THEOREM AND UNDER THE THEOREM WHEN ADAPTIVE EXPECTATIONS ARE ASSUMED.

IV. AN ALTERNATIVE BUT EQUIVALENT FORMULATION

It is possible to develop an alternative model based on the traditional Marshallian distinction between long and short runs which leads to precisely the same conclusions reached in the preceding section. The usual distinction between long-run and short-run elasticities of supply has generally been made on the basis of the elasticities of supply of certain factors to the firm.⁹ The contention is that in the shortest of short runs most or all factors of production are fixed while, as time passes, successively more of these restrictions are removed. Under this interpretation of the meaning of different runs it can be shown that the short-run elasticity of supply for an individual

9. Milton Friedman, "Lectures in Price Theory," unpublished mimeographed notes (Chicago: 1951), pp. 119-24.

firm is always less than or equal to the long-run elasticity of supply. The argument hinges on the fact that the minimum cost of producing an additional unit of output cannot be less when an entrepreneur's freedom of choice is limited than when he is completely free to choose, because he always has the alternative of choosing exactly that situation to which he was previously limited.¹ Thus, from any point on a long-run supply schedule, we may think of a fan of short-run supply curves which gradually approach the long-run supply schedule, from the right above it, and from the left below it. Åkerman has pictured this situation graphically.²

In order to develop a mathematical model of cobweb phenomena based on the considerations discussed above, it is not necessary to reject Ezekiel's original assumption that farmers react to the price at the preceding harvest. Although it is possible to retain the assumption about expectations of long-run normal price outlined in Section III, this will not be done here. Instead, we assume that supply is a function of last period's price, P_{t-1} .

The considerations of the first paragraph of this section suggest the following simple model of the relation between current output and long-run equilibrium output; Let \bar{q}_t^S be the long-run equilibrium quantity supplied, in contrast to q_t^S which is the current quantity supplied. Let $0 < \gamma_s \leq 1$ be a constant of proportionality which we shall call the coefficient of adjustment. The following equation, when coupled with a long-run supply curve, yields a fan of short-run supply curves through each point of the long-run curve:

$$(17) \quad q_t^S - q_{t-1}^S = \gamma_s [\bar{q}_t^S - q_{t-1}^S], \quad 0 < \gamma_s \leq 1.$$

In words, the rate of adjustment to long-run equilibrium is proportional to the difference between current output and the long-run equilibrium output.³

Let the demand curve be equation (6) as before, but now let the long-run supply curve be

$$(18) \quad \bar{q}_t^S = c + dP_{t-1}.$$

That is, suppose farmers base their long-term plans on price the preceding harvest, but that they cannot or do not realize these

1. *Ibid.*, p. 122.

2. Åkerman, *op. cit.*, Figure I.

3. The characteristics of this model are discussed more fully in *Distributed Lags and Demand Analysis*, *op. cit.*, and in "Estimates of the Elasticities of Supply of Corn, Cotton, and Wheat," *op. cit.*, pp. 59-67. I have used it in quite a different context in my "A Note on the Long-Run Automobile Demand," *Journal of Marketing*, XXII (July 1957), 57-64.

plans in the short run. We assume market equilibrium in both the long and short runs; hence,

$$(19) \quad \bar{q}_t^S = \bar{q}_t^D = \bar{q}_t \quad \text{and} \\ q_t^S = q_t^D = q_t.$$

Since no difference between short- and long-run demand is assumed, the current and long-run quantities demanded are one and the same. By a procedure similar to that used in the preceding section, we have from equations (6) and (17)–(19):

$$(20) \quad P_t - \left[\left(\frac{d}{b} - 1 \right) \gamma_s + 1 \right] P_{t-1} = \frac{(c - a)\gamma_s}{b}.$$

Equation (20) is precisely the difference equation (11) with γ_s in place of β . It therefore has the same solution, and the necessary and sufficient conditions for a convergent cobweb are

$$(21) \quad 1 - \frac{2}{\gamma_s} < \frac{d}{b} < 1.$$

The model expressed by equations (6), (17), and (18) reduces to the ordinary cobweb model when $\gamma_s = 1$.

V. SOME EMPIRICAL RESULTS

In previous sections we have developed two models of cobweb phenomena which, although based on different considerations, lead to the same necessary and sufficient condition for stability. In this section we present empirical estimates of the long-run elasticities of supply of corn, cotton, and wheat based on models of the types discussed above. These are combined with existing estimates of the elasticities of demand to yield measures of the ratio d/b for the three commodities.⁴ These ratios are then compared with the lower limit $1 - 2/\beta$ or $1 - 2/\gamma_s$.

Previous estimates of the elasticities of demand for corn, cotton, and wheat at the farm level indicate that they range from a low of about $-.1$ for wheat to about $-.3$ for cotton and a high of $-.6$ for corn.⁵

Estimates of the elasticities of supply of corn, cotton, and wheat have previously been derived on the same general assumptions as

4. In the neighborhood of equilibrium, the ratio of the elasticity of supply to the elasticity of demand crudely measures the ratio of the two slopes.

5. See Richard J. Foote, John W. Klein, and Malcolm Clough, *The Demand and Price Structure for Corn and Total Feed Concentrates*, U. S. Department of Agriculture Technical Bulletin No. 1061 (1952), pp. 39–41; Frank Lowenstein and Martin S. Simon, "Analyses of Factors that Affect Mill Consumption of Cotton in the United States," *Agricultural Economics Research*, VI (1954), 101–10; and Kenneth W. Meinken, "The Demand and Price Structure for Wheat," U. S. Department of Agriculture Technical Bulletin No. 1136 (1955), pp. 42–43.

those on which the models of sections III and IV are based.⁶ The estimates of the elasticities of supply and β or γ_s have been derived by two methods: one I have called the "iterative" method, and the other the "noniterative." For an explanation of the two methods the reader is referred to the sources of the estimates. The period of estimation was 1909-32. The estimates obtained by the iterative method range from a low of 0.4 for corn to 1.2 for wheat and a high of 4.5 for cotton; those obtained by the noniterative range from a low of 0.2 for corn to 0.7 for cotton and a high of 0.9 for wheat. As the two sets of estimates differ considerably, it is well to regard the results obtained with either set as essentially illustrative.

As a by-product of the estimates of the long-run elasticities of supply described in the preceding paragraph, estimates of the coefficients of expectations, β , or the coefficient of adjustment, γ_s , were derived. The two models presented above suggest precisely the same form of regression, so that what is called β in one regression may equally well be interpreted as γ_s . Which interpretation we use makes little difference in examining the results for stability. The iterative estimates of β or γ_s range from a low of 0.04 for cotton to 0.25 for corn and a high of 0.4 for wheat; the noniterative estimates are 0.4 for cotton and 0.5 for wheat and corn.

Table I gives the ratios of the elasticity of supply to the elasticity of demand for cotton, wheat, and corn for the two methods of estimation (columns (3) and (4)). Columns (5) and (6) present the

TABLE I

COTTON, WHEAT, AND CORN: STABILITY OF AN EQUILIBRIUM OF DEMAND AND SUPPLY FOR VARIOUS ESTIMATES OF THE ELASTICITY OF SUPPLY AND THE COEFFICIENTS OF EXPECTATIONS OR ADJUSTMENT

Crop	Assumed Elasticity of Demand ¹	Estimated Elasticity of Supply ÷ Assumed Elasticity of Demand ²		Estimated $(1-2/\beta)$ or $(1-2/\gamma_s)$ ²	
		Iterative Estimate of the Elasticity of Supply	Noniterative Estimate of the Elasticity of Supply	Iterative Estimate	Noniterative Estimate
(1)	(2)	(3)	(4)	(5)	(6)
Cotton	-.3	-15.1*	-2.2*	-49.0	-3.9
Wheat	-.1	-11.8	-9.3	- 4.4	-2.8
Corn	-.6	- .7*	- .3*	- 7.0	-2.7

* Stable equilibrium indicated.

1. Sources cited in footnote 5, p. 236.

2. Sources cited in footnote 6 below.

6. See "Estimates of the Elasticities of Supply of Selected Agricultural Commodities," *op. cit.*, and "Estimates of the Elasticities of Supply of Corn, Cotton, and Wheat," *op. cit.*, pp. 247-96.

values of $1-2/\beta$ or $1-2/\gamma_s$ for the two methods and the three crops. A comparison of one of the columns (3) or (4) with its corresponding member of the two columns (5) and (6) indicates the nature of the equilibrium between demand and supply. If the ratio in column (3) or (4) is less in absolute value than the corresponding ratio in column (5) or (6), then the equilibrium is stable. If it is greater in absolute value, the equilibrium is unstable. If the two are equal, the equilibrium is neutral in the sense that perpetual but nonexplosive fluctuations occur if the equilibrium is disturbed. Ratios indicating a stable equilibrium are marked with an asterisk. Only wheat shows indication that its equilibrium is unstable, and this holds regardless of which method we use to estimate the elasticity of supply. The other crops, cotton and corn, appear to have stable equilibriums of demand and supply.

It may seem somewhat difficult to believe that the equilibrium of demand and supply is actually unstable in the case of wheat. We should, however, remember that instability may exist only within a certain range of prices; while the demand for wheat may, in fact, be highly inelastic in the range of prices which prevailed during the period used in estimation, it is probable that at a lower range of prices it is highly elastic. At a somewhat lower range of prices, wheat would be used for feed and would compete strongly with corn. An elasticity of demand for wheat of only $-.3$ or $-.35$ would suffice to turn an indication of instability into one of stability; at higher prices the elasticity of supply may be less; hence, instability may occur only within a relatively narrow range around equilibrium.

VI. MORE COMPLICATED MODELS

It is possible to develop more complex models of cobweb phenomena by techniques similar to those used in preceding sections. As an indication of how to proceed, one such model is presented in this section; it is not, however, fully analyzed.

Let \bar{q}_t^D be the long-run equilibrium quantity demanded and q_t^D be the current quantity. Let \bar{q}_t^S and q_t^S be, as before, the long-run equilibrium quantity supplied and the current quantity supplied, respectively. Our model is specified by six equations: The long-run demand equation

$$(22) \quad \bar{q}_t^D = a + bP_t ;$$

the demand adjustment equation

$$(23) \quad q_t^D - q_{t-1}^D = \gamma_D [\bar{q}_t^D - q_{t-1}^D], \quad 0 < \gamma_D \leq 1 ;^7$$

7. Equations (22) and (23) taken together imply that the current quantity demanded depends on the price taken with a distributed lag. Under the present

the long-run supply equation

$$(24) \quad \bar{q}_t^S = c + dP_{t-1};$$

the supply adjustment equation

$$(25) \quad q_t^S - q_{t-1}^S = \gamma_s [\bar{q}_t^S - q_{t-1}^S], \quad 0 < \gamma_s \leq 1;$$

and, finally, two market equilibrium conditions, short-run

$$(26) \quad q_t^D = q_t^S = q_t,$$

and long-run

$$(27) \quad \bar{q}_t^D = \bar{q}_t^S = \bar{q}_t.$$

A second-order difference equation may be derived from equations (22)–(27) as follows: From (22), (23), and (26) and (27) we have

$$(28) \quad q_t = a\gamma_D + b\gamma_D P_t + (1 - \gamma_D)q_{t-1},$$

and from (24), (25), (26), and (27) we have

$$(29) \quad q_t = c\gamma_s + d\gamma_s P_{t-1} + (1 - \gamma_s)q_{t-1}.$$

$$(30) \quad b\gamma_D P_t - d\gamma_s P_{t-1} + a\gamma_D - c\gamma_s = (\gamma_D - \gamma_s)q_{t-1};$$

eliminating q_{t-1}

$$(31) \quad a\gamma_D(1 - \gamma_s) - c\gamma_s(1 - \gamma_D) + b\gamma_d(1 - \gamma_s)P_t - d\gamma_s(1 - \gamma_D)P_{t-1} \\ = (\gamma_D - \gamma_s)q_t.$$

Lagging (31) by one period and substituting the result in (30), we have a second-order difference equation.

$$(32) \quad P_t - \left[\left(\frac{d}{b\gamma_D} - 1 \right) \gamma_s + 1 \right] P_{t-1} + \frac{d}{b} \frac{\gamma_s}{\gamma_D} (1 - \gamma_D) P_{t-2} = \\ \frac{(c-a)\gamma_s}{b}.$$

This difference equation reduces to equation (20) in section IV when $\gamma_D = 1$. It may be solved but its solution is not presented here.

Other more complicated models involving stocks and/or expectations could be constructed, but the example presented above should suffice to indicate the possibilities.

interpretation this means that the long-run elasticity of demand differs from and is greater than the short-run elasticity. That this should be so becomes reasonable when we consider that our demand includes an excess supply from stocks.

VII. CONCLUSIONS

It is possible to construct more sophisticated models of cobweb phenomena than the model which was first presented by Ezekiel in the pages of this *Journal*. In particular, mathematical models may be constructed which distinguish between long- and short-run elasticities of supply and which incorporate both a long-run supply curve and its associated short-run supply curves into the analytical framework. These models are more suggestive of, and more amenable to, empirical application than Åkerman's recent diagrammatic extension of the cobweb theorem.

The models presented in this article show that, while the range of possible instability is lessened when account is taken of the distinction between long- and short-run supply schedules, the possibility still exists. The illustrative empirical results which are presented suggest that wheat, in the absence of price support programs, may be characterized by such instability, at least in the neighborhood of equilibrium.

The methods outlined here may be used to construct further models of cobweb phenomena which are easily adapted to fit complex, individual situations which we may wish to investigate.

MARC NERLOVE.

U. S. AGRICULTURAL MARKETING SERVICE

COMPETITION, CREDIT POLICIES, AND THE CAPTIVE FINANCE COMPANY

By PAUL H. BANNER

Introduction, 241. — The growth of captive finance companies, 242. — Effects on competition, 243. — Other examples of the competitive use of credit, 246. — Noncompetitive uses of credit, 248. — Effects on credit policy, 248. — Individual industry experience, 251. — Summary, 257.

The development of the captive finance company¹ — the finance company wholly owned and operated as a subsidiary of a nonfinancial firm — presents new problems in the field of consumer credit. The captive finance company has as its primary purpose the underwriting of the sale of the parent company's products and it is therefore influenced in its actions and policies by merchandising requirements, not by objective monetary standards and credit conditions. It is the purpose of this study to determine the effect of this sales device upon competition and upon credit conditions. This study suggests that the captive finance firm is most effectively used by, and is a competitive advantage primarily available to, the large company. A second conclusion is that consumer credit controls now available to monetary authorities are largely ineffective in dealing with this rising source of consumer credit. It is possible therefore that either direct controls will be required to regulate consumer installment debt expansion or divestiture will become necessary if the current growth of captive finance companies continues.

Installment credit is an absolute necessity in certain industries. The automobile industry, for example, could not have grown to its present size without some form of time payment plan. Credit, however, is not uniformly available as a sales method. Inadequacy of credit facilities has been a chronic complaint in industrial and agricultural history. In part, this complaint stems from the fact that credit institutions normally approach new uses of credit cautiously. New industries present greater risks than old ones. The prefabricated house and home trailer industries are current examples — the auto-

1. This study is concerned only with the financing of sales by nonfinancial firms, in most cases manufacturing concerns. Factoring is not included since it does not present the same competitive and monetary problems. Ownership of financing affiliates by other financial firms is also excluded from consideration since this is a problem of concentration and no new monetary problems are at issue.

mobile industry an illustration of a few decades ago. Inadequacy of credit facilities is a relative matter, however, and in many industries today certain firms feel this inadequacy because of their inability to meet the credit terms offered by their competitors operating captive finance firms.

THE GROWTH OF CAPTIVE FINANCE COMPANIES

This study indicates that the number of captive finance companies has been increasing continuously. Since 1948 their number has doubled and in recent years there has been an acceleration in their formation.

ORGANIZATION OF CAPTIVE FINANCE COMPANIES¹

Year of Organization	Number Organized
1948 and prior years	19
1949	2
1950	3
1951	5
1952	1
1953	3
1954	5
1955	5
1956	7
10 months 1957	8
Unknown	11
Total	69
Number of parent firms	55

1. Sources utilized in the preparation of this study include: "A List of 1000 Large Manufacturing Companies, Their Subsidiaries and Affiliates," Federal Trade Commission, 1951; *Standard and Poors Corp. Records*; *Moody's Industrial Manual*; *Wall Street Journal*; *Journal of Commerce*. Company reports and stock registration statements were also used. The above summary does not represent a complete census of financing affiliates. There is no directory of finance companies reporting data such as is required to determine ownership. Subsidiary corporations are not always reported separately in financial manuals or they are referred to so obliquely as to make their operations difficult to determine. In certain cases in annual reports partially owned firms are treated as investments rather than as subsidiaries and little or no information is offered as to their operations. If the parent firm is not registered on an exchange, difficulties are compounded.

Most of the captive finance firms studied are affiliates of firms listed as among the 500 largest industrial firms in the country by *Fortune* magazine in its July 1956 Supplement. To some extent this indicates a limitation on the quality of the sample. It is known, for example, that automobile dealers frequently operate finance companies in conjunction with dealerships. If the number of firms with finance companies is considered small, however, it should be noted that many industry products are of such low unit cost that financing of sales is unnecessary. Also, in those cases where industry products are intermediary rather than consumer products, different forms of financing are utilized. In the petroleum, chemical, food and textile

industries, for example, sales financing is unnecessary to move a significant proportion of industry product. Among the top 25 firms ranked according to sales are 8 oil firms, 3 food processors and 2 chemical firms, all of which would not be expected to utilize installment sales techniques.

EFFECTS ON COMPETITION

Frequently it has been found that ability to provide credit is a more effective competitive weapon than an attractive product price. Many instances, especially in automobiles and housing, can be found where the absolute level of monthly payment, not the over-all cost of the product, is the determinant of sale. Flexibility in financing arrangements can be more important than total interest cost in selling a product. A captive finance firm whose only purpose is to further the sale of the parent firm's products can, therefore, provide competitive advantage.

Examples of the advantage to a producer in having a financing affiliate are found in many industries. Bus producers have insisted that financing is one of the chief advantages of General Motors. ACF Brill's president, for example, stated that his buses were priced competitively but because he could not offer financing for so long a period, he lost many sales.² The President of the Southern Coach Manufacturing Company offered similar evidence. His firm was specifically told it could have sales if it could provide financing. His firm now makes no attempt to get customers unless it knows such customers can finance purchases locally.³ A representative of another firm, the Flxible Company, stated that his firm had utilized General Motors Acceptance Corporation (GMAC) to finance its sales. He reported an instance of GMAC refusing to handle his financing of a bus sale and General Motors thereupon receiving the same order which it financed on longer terms through Yellow Motors Acceptance Corporation, another General Motors affiliate.⁴ Flxible now relies on customer financing by local banks.⁵ The Ford Motor Company

2. *Hearings*, U. S. Senate, Committee on the Judiciary, Subcommittee on Antitrust and Monopoly, 84th Congress, 1st Session (1955), Vol. VI, p. 2655, hereafter referred to as *Hearings*.

3. *Hearings*, p. 2659.

4. General Motors Acceptance Corporation (GMAC) and Yellow Motors Acceptance Corporation (YMAC) are wholly-owned subsidiaries of General Motors Corporation. GMAC finances sales of the passenger cars and appliance divisions of GM while YMAC's principal business is the wholesale and retail financing of trucks and coaches manufactured by the GMC Truck and Coach Division of GM, Euclid Division products and trailers, bodies, and special equipment sold in conjunction with trucks, coaches and Euclid road moving machines.

5. *Hearings*, p. 2665.

reportedly left the bus business not only because the market was declining but also because financing of customers was necessary.

On July 6, 1956, the Department of Justice filed an antitrust suit against General Motors charging among other things monopolization of the manufacture of buses and the financing of the sale of buses through YMAC on terms which General Motors' competitors with more limited resources could not meet. One element of relief requested was that General Motors be required to offer to finance the sales of buses manufactured by any other company upon the same terms and conditions as it finances its own buses.⁶

Probably in no industry has credit become so important a sales device as in the retailing of automobiles. Only one firm, General Motors, owns a sales finance company. The problem posed will be discussed hereafter as part of a general discussion of the automobile industry, but at this time it can be pointed out that in 1954-55 there occurred a major challenge to General Motors' leadership in the automobile industry. The Ford Motor Company, planning a public offering of its stock, drove the industry into the most competitive battle for sales leadership in the postwar period. One retaliatory device resorted to by General Motors was liberalization of credit. This liberalization occurred only shortly after the firm in a policy announcement declared:

"GMAC's current experience shows that where down-payments on new cars are less than one-third, reposessions are disproportionately higher, as dealers are well aware."⁷

Shortly after this statement GMAC's policies were changed and the number of transactions involving down payments lower than one-third about doubled percentagewise.⁸ This flexibility had an immediate effect upon dealers of other cars and upon competing finance companies. A dealer testifying before the Subcommittee on Antitrust and Monopoly stated that the finance firm he had used had always offered competitive rates in the past, but in early 1955 he found he was losing sales to GM dealers because of financing. GMAC was taking automobile paper that in the past it would not have taken.⁹

There are insufficient data available to reveal whether GMAC

6. *United States v. General Motors Corporation*, Civil Action #15816, July 6, 1956.

7. *Hearings*, p. 3999.

8. *A Study of the Antitrust Laws*, Staff Report of the Subcommittee on Antitrust and Monopoly, Committee on the Judiciary, U.S. Senate, 84th Congress, 1st Session, p. 70.

9. *Hearings*, p. 3104.

initiated the liberalization of automobile credit terms in 1954 and 1955 which resulted in the greatest increase in automobile credit in any single year, but from 1954 to 1955 it can be shown that the major finance firms, the most important of which is GMAC, were more liberal than their smaller competitors in the granting of credit. In 1954 the median credit on new car purchases for the four major companies in this field was \$1,590 in comparison with \$1,650 for all other sales finance companies. In 1955 the median credit allowed rose to \$1,790 and \$1,770, respectively, for the four major companies increased the median size of loans \$200 in one year, while their smaller competitors increased the median size of their loans only \$120.¹ General Motors' share of the passenger car market in this booming credit market rose to its highest percentage in history. In a market in which credit purchases rose by one-half while cash purchases rose less than one-sixth, the firm with a financing affiliate found its market position considerably enhanced.

Credit liberalization is so effective as a competitive weapon in this industry because of the high number of sales involving trade-ins. More than four of every five new car sales involve trade-ins so that

SELECTED AVERAGES FOR CREDIT PURCHASES

(Mean dollar amounts)

Item	1955	1954
Contract car price	\$2890	\$2690
Effective car price ¹	2480	2320
Value of car traded-in or sold	590	580
Net outlay	1900	1740
Amount of additional cash ²	210	240
Amount of credit ³	1690	1500

Source: Board of Governors, Federal Reserve System, *Consumer Instalment Credit*, Part IV, "Financing New Car Purchases," 1957, p. 65.

1. Contract car price less estimated overallowance on trade-in, if any.

2. Cash paid less encumbrance on trade-in, if any.

3. Excludes finance and insurance charges; excludes any supplementary borrowing for down payment.

cash requirements for the purchase of a new car represent the difference between down-payment requirement and the trade-in value of the used car. Liberalization of credit can reduce cash requirements to a negligible amount. A reduction in down-payment requirements from one-third to one-fourth on a \$2,400 car is \$200. A three-year-old car which cost when new \$2,400, is roughly worth \$800, or the one-third down payment on a new car without any further cash

1. "Financing New Car Purchases," *Consumer Instalment Credit*, Part IV, Board of Governors of the Federal Reserve System, 1957, p. 73.

payment. A four-year-old car would be worth about \$650, or would require an additional \$150 cash for a new car purchase on credit. A decline in down-payment requirements, however, to one-fourth or \$600, would mean that another age group of cars could be tapped as a market for new car sales without cash down payment. According to Federal Reserve Board data the median interval before repurchase of a car declined from 37 months in 1954 to 32 months in 1955 for credit buyers of automobiles.² Concomitant with this decline in the median time interval on transactions, credit rose and cash paid declined.

While these important changes in the character of automobile installment paper occurred one item did not change, the average monthly payment on new car installment contracts. This remained constant from 1954 to 1955 at \$72. Thus the lower down payments were accompanied by a lengthening of maturities, the median for which rose from 24 months in 1954 to 30 months in 1955.³

The effect can be seen in the age distribution of cars traded-in on purchases of new cars. In 1953, 50 per cent of all cars traded-in on new cars were three years old or less. Cars four and five years old comprised 20 per cent of cars traded-in. In 1954 and 1955, cars aged three years old or less declined to 45 per cent and 42 per cent, respectively, while cars four and five years old rose to 28 per cent in 1955. Thus concomitant with a decline in down-payment requirements the importance of four- and five-year-old cars as trade-ins rose 50 per cent.

OTHER EXAMPLES OF THE COMPETITIVE USE OF CREDIT

Another competitive use of the finance company is in the introduction of a new product. The firm offering liberal financing makes its sale easier. A recent example is the automatic pinsetter for bowling alleys. The substitution of capital equipment for labor in bowling alleys is a large investment for a business which has not previously required high capital outlay. The Murray Corporation and Brunswick-Balk-Collender Company, developers and marketers of this product, formed an affiliate, the Brunswick Murray Automatic Pinsetter Corporation, to finance installment sales of these machines. Competition was undoubtedly a factor also. A rival product is offered on a rental basis by its developer.

An older example of the use of credit to introduce a product is the diesel locomotive. When General Motors introduced its diesel engines 40 per cent of the mileage of American railroads was in

2. *Consumer Instalment Credit*, *op. cit.*, p. 37.

3. *Consumer Instalment Credit*, *op. cit.*, p. 37.

receivership. Obviously railroad credit was poor, and in the depression period capital expenditures were avoided. In order to introduce its product General Motors offered to finance its sale with no down payment required. A revolving credit of \$5 million was established and engines were leased to railroads with purchase options. In 1937 and 1938, GMAC financed 12 per cent and 67 per cent, respectively, of GM sales to domestic railroads.⁴

In the appliance industry when production declined in 1956 and various firms dropped lines or sold out completely, General Electric turned to easier consumer financing as a stimulus to sales. It announced that a home owner, or long-term renter, who made structural changes in his kitchen or laundry could have the whole project financed through General Electric Credit Corporation, if at least two major appliances were bought. The President of the Philco Corporation, a firm with only one-tenth of the sales of GE, commented "GE's new financing plan will have a lot to do with sales and will make it very rough for other manufacturers. Only a big outfit like GE could do such a thing."⁵ The terms offered by GE include 10 per cent down payment with five years to pay the balance. Philco is a well-known name in the appliance industry. The effect of GE's competition on smaller firms can be imagined.

Financing terms are powerful competitive tools which can best be utilized by the large firm which has better access to the money market and can borrow at lower interest rates. The large firm invariably has established sources of financing. Its name is known and it has established credit status. This is of importance since capitalization of a finance firm is less important than its borrowing ability both in the long-term capital market and in that for short-term money. With the parent firm's credit status, name, and reputation, the newly formed financing company has little difficulty in securing entry into the capital market. Sears, Roebuck, for example, upon the organization of its financing affiliate, did not find it necessary to operate its subsidiary using capital funds alone until it had built up a credit rating. Immediately upon the organization of the subsidiary in November 1956 with a capitalization of \$35 million, a \$50 million debenture issue was floated, underwritten by leading security houses and repeated soon thereafter. The firm which handled the greatest share of the issue was the firm which in the ordinary course of Sears' business handled its financing and the sale of its commercial paper.

4. *Hearings, op. cit.* VIII, 4352.

5. Raymond A. Rich, Philco Corporation, quoted in *Wall Street Journal*, June 4, 1956.

With reference to the cost of money, a major part of the finance company's cost of doing business, one need only consult the difference in interest rates according to size of borrower. Rates normally vary with the size of loan. In addition, the firm which can place its short-term paper directly secures advantageous rates in comparison with paper placed through brokers. The average money market rate in 1956, for example, was 3.06 per cent for directly placed finance company paper in comparison with a rate of 3.31 per cent for prime commercial paper of comparable maturity.⁶ Thus it is not only that the large firm can more readily utilize financing terms as competitive weapons, but the advantages of the parent firm's size can be translated into advantages of size in the money market.

NONCOMPETITIVE USES OF CREDIT

Interesting variations in sales financing arrangements are occurring in the carpet and trailer truck industries, variations which would appear to benefit the smaller rather than the bigger firms in the industry.

Sales financing arrangements in the carpet industry have developed in such a manner as to minimize competitive effects among carpet producers. In September 1955 Mohawk Carpet Mills arranged with Allied Building Credits, a subsidiary of the Transamerica Corporation, to finance dealer sales of its products. In succeeding months other firms made similar arrangements with the same finance company. In October 1956 Alexander Smith, Firth Carpet Company, A and M Karghensian and C. H. Marland and Sons simultaneously announced these arrangements and in December James Lees and Sons. Other mills in the country are reported to have done likewise. Terms offered are similar to all users and provide no advantage to any particular manufacturer. Such an arrangement is apparently the only noncompetitive intra-industry credit arrangement in effect.

Another noncompetitive credit arrangement different in purpose has been suggested for the truck trailer industry. In this industry a credit pool has been proposed to help both small truckers and producers.

EFFECTS ON CREDIT POLICY

The existence of large captive finance companies owned and controlled by leading manufacturers in their respective fields cannot but affect the credit structure of the country. The captive finance company is not an independent business entity. The fundamental objective of the company is not necessarily the same as that of an inde-

6. *Federal Reserve Bulletin*, Jan. 1957, p. 53.

pendent finance company. Is it to function principally as a finance company, or is it to function as a means of furthering the interest of the parent? It would appear that the paramount purpose is usually the latter. If so, then the existence of captive finance companies, as major suppliers of consumer credit, can create conflict with over-all national credit policy. It may be to the competitive advantage of the parent to adopt policies at variance with monetary objectives and to be in a position to do so. If the government, for example, is fostering credit restriction to limit business borrowings for inventory purposes, banks will more carefully review consumer credit applications and generally slow down their lending activities. If captive finance firms are the sources of credit for dealer inventories, pressures from the parent firm to continue lending may be too strong to resist even though further lending may be unprofitable. Independent market forces do not directly pass judgment on this demand for credit. One automobile dealer informed the author that his greatest protection from factory pressure to buy more cars was the independent finance company he used, which exercised independent judgment as to how many cars it would floor plan for him. He obviously did not build up the large inventories many other dealers did in the latter part of 1955.

Structurally, there are other effects of the captive finance company. In the market for funds, the captive finance company probably has better credit than most independent finance companies. Its affiliation with an important manufacturing firm is advantageous. It has the credit of the parent organization behind it. It is therefore a preferred credit risk to the independent finance company. Credit policies that reduce the availability of credit will bear more heavily upon the independent than upon the captive firm of comparable size. The captive firm, though smaller than many independent firms, as an affiliate of a large manufacturing firm with nationwide sales, can secure credit lines in many banks.

Complementary to availability of credit as a limit upon the activity of finance companies in a tight market is the treatment of credit risks. With higher rates and restricted availability, independent firms survey their borrowers more carefully, ration credit, and react in the manner monetary authorities anticipate. Captive finance companies cannot be relied upon to react in this manner. It must always be remembered that the captive firm exists primarily to promote the sale of the parent firm's products. Not only may it not feel the usual profit incentives operating upon the independent firm, but knowing that these conditions exist it may take advantage of

the situation to enhance the parent firm's market position. If the captive firm is to be used to further the parent firm's market position, it cannot allow itself to be limited by the same criteria as limit the independent firm.

The captive firm usually has much better knowledge of its customers' positions than a banking relationship alone would supply. It surveys its own industry and while perhaps more parochial in outlook, it is more knowledgeable of its specialized field. Hesitation in judgment of credit worthiness or a general lagging in the granting of credit is unlikely.

More important, however, is the positive use of credit. If credit costs are rising it may be advantageous to absorb this higher cost and if necessary operate at a deficit to enhance a market position. The very limitation on credit for competitors' products may be the impetus to provide credit for the parent firm's products. At such a time new markets can be invaded and a broader distribution of products achieved. These are advantages to the firm which would impel it to act counter to current credit trends.

The above discussion has been applicable to a period of restricted credit. What of a period of liberal credit? In such instances either the captive firm can retire from the market and act passively or it can utilize its preferred position with the parent firm's customers to assure itself of a market for its services. This has been done. Pressure to use the financing and other services offered by a parent firm are not uncommon.

An example of credit policies entirely subservient to the interests of the parent corporation was previously noted in the action GMAC took in 1954-55. The same flexibility in interest rates was noted in the case of General Electric which in 1956 announced a liberalization of financing terms in a period of continuously rising interest rates and attempts by monetary authorities to limit availability of funds for credit expansion. The president of General Electric at a press conference praised the Federal Reserve Board's policy of credit restraint but at the same time announced that the GE credit company was helping those feeling the pinch.⁷

The tight money market has not prevented the growth of the captive finance company. It is possible that there is no casual relationship and the growth can be ascribed to demonstration effect. On the other hand, the very tightness of the market may be contributory and in addition the profitability of installment paper, in comparison with shrinking profit margins, may be influential. Large firms, and

7. Ralph J. Cordiner, quoted in *Wall Street Journal*, Dec. 14, 1956.

those studied were primarily among the largest manufacturing firms in the country, usually have established sources of credit and can be expected to attempt to circumvent credit restrictions as a barrier to sales. Furthermore, the large firm with a distributive organization can readily expand its services to include sales financing. Ease of entry is not a problem. The effect of such entry is a problem.

INDIVIDUAL INDUSTRY EXPERIENCE

The captive finance firm is usually a specialized firm. It handles paper arising from the operations of the parent corporation, and rarely handles other financing.⁸ There can be, therefore, significant differences among firms, and no uniform pattern of operations can be expected. Credit requirements, procedures and practices vary among industries and the captive firm will be designed to fit industry requirements. Indeed the very existence of the firm may be in part due to the inadequacy, or lack of flexibility of alternative financing institutions. Channels of distribution, seasonality, and historic practices vary among industries and will influence credit requirements.

The motor vehicle industry is particularly dependent upon installment credit. Approximately two-thirds of all passenger car sales involve the use of credit and automobile paper accounts for almost one-half of the total outstanding installment credit. In an industry in which credit is of such importance, only General Motors owns its own finance company, and until recently its two major competitors were restricted by law from any affiliation with such a firm. In 1938 both Chrysler and Ford were prohibited by court order as a result of an antitrust action from any such affiliation, an order which was modified by the Supreme Court in 1948. General Motors now owns and operates two finance companies, General Motors Acceptance Corporation, the largest sales finance company in the world, and Yellow Motors Acceptance Corporation, a separate firm to finance the sale of GMC trucks and other heavy equipment.

Credit has always been a problem in the automobile industry, and in the twenties considerable attention was devoted by automobile firms to methods of supplementing credit facilities. While the automobile industry was profitable, there were many failures and the industry credit was low. The need for supplemental financing in the automobile industry arose from the unwillingness of banks to finance the installment sale of automobiles. Specialized lending facilities in

8. General Motors, for example, financed the sale of 752,506 new automobiles in 1954, of which only 4,027 were non-GM products, or only one-half of 1 per cent of its retail auto financing.

the forms of sales finance firms grew rapidly after World War I, and without the ethical practices of the banking profession. Exorbitant finance charges, excessive insurance charges, "packs" fees and so forth were loaded upon the auto purchaser, and the charges quoted in a purposely ambiguous manner so as to confuse the buyer. These practices, while still utilized today, were more blatant in the twenties when alternative financing sources were unavailable. The excessive charges for automobile financing detracted from the reputation of the industry and in addition raised the price of the product to the consumer. Automobile firms, therefore, defensively concerned themselves with sales financing and sources of credit.

General Motors organized its financing affiliate in 1919. Ford organized the Universal Credit Corporation in 1928. Other firms made agreements with "factory preferred" finance companies. Arrangements in the latter case varied but generally in return for lending at agreed upon rates, the automobile firm paid a subsidy to the finance firm or assumed responsibility for repossessions.

General Motors initially provided its financing services as supplementary financing, but this practice was soon modified. GMAC financed on a recourse basis, that is, the dealer was obligated to buy back repossessed cars, a risk dealers preferred to avoid. Furthermore, it was profitable for the dealer to enter an agreement with a local finance company. In 1925 GMAC wooed dealers by offering a rebate on finance charges in the form of a dealers' reserve equal to 20 per cent of the finance charge on new car sales and 30 per cent on used car sales. Simultaneously pressure was exerted upon dealers to use the affiliate's services. By 1935, when it was estimated that 55 per cent of all new cars were sold on installment sales plans,⁹ GMAC's retail financing equalled 46 per cent of General Motors car sales, which would imply that GMAC secured about 80 per cent of its dealers' finance business.

In 1935 General Motors announced its 6 per cent plan which provided the purchaser with a quick method of computing his total finance charges. The plan while not the same as 6 per cent simple interest offered lower interest rates than were currently being charged by other finance companies and forced a competitive reduction in finance charges. The competitive advantage of the plan can be seen in that from 1935 to 1936 GMAC retail financing increased 50 per cent while car sales of General Motors increased 21 per cent in value.

9. National Association of Sales Finance Companies, quoted in W. C. Plummer and R. A. Young, *Sales Finance Companies and Their Audit Practices*, National Bureau of Economic Research (New York, 1940), p. 95.

General Motors' market share in the passenger car field also rose, from 38 per cent in 1935 to 43 per cent in 1936. In the following year as other firms met the GMAC rate, General Motors' market share dropped back.

The importance General Motors attached to financing can also be illustrated by its action in Germany upon purchasing an 80 per cent interest in Adam Opel A.G. When it purchased this firm in 1929 it immediately formed a Germany Acceptance subsidiary. Opel Finanzierungs G.m.b.H.¹

Chrysler Corporation had an agreement with Commercial Credit Corporation, an independent finance company, from the time of the former's reorganization. In 1935 this agreement was revised and Chrysler bought considerable stock in the finance company. The agreement provided that Commercial Credit pay Chrysler a percentage of its earnings and that Commercial Credit provide financing terms for Chrysler dealers as favorable as those given to all but Ford dealers. Where preferable terms were given Ford dealers, their names were supplied the Chrysler Corporation. Chrysler could request Commercial Credit to provide lower rates but would have to pay the subsequent difference in interest rates. Chrysler dealer and distributor contracts were revised to include the provision that charges in excess of those of Commercial Credit could not be charged. Pressure was also applied upon dealers to use Commercial Credit financing. In return Chrysler profited and in the three years this contract ran, Commercial Credit paid Chrysler more than \$3 million.²

Ford organized the Universal Credit Corporation in 1928 to finance the sale of its cars. In 1933 Ford sold Universal to CIT. Ford also owned a finance company in Japan, the Fordson Finance Company, but it is unknown whether this firm is still in existence.

On May 27, 1938 the Department of Justice, as a result of complaints of independent finance companies and automobile dealers, instituted an antitrust suit against the big three automobile producers charging a conspiracy in restraint of interstate commerce and coercion of dealers to use factory affiliated finance companies. On November 15, 1938, civil suits were substituted for criminal indictments and consent judgments entered in the cases of Ford and Chrysler. General Motors preferred to stand trial. The court's verdict held General Motors guilty and fines were imposed. General

1. *Annual Report*, General Motors Corporation, 1929.

2. *Federal Trade Commission Report on Motor Vehicle Industry*, 76th Congress, 1st Session, H.Doc. No. 468, p. 614.

Motors appealed and the judgment was affirmed in 1941.³ A civil suit was thereupon filed seeking divestiture. The final decree against General Motors was entered in 1952 and was similar to the judgments entered against Chrysler and Ford. In each case coercion of dealers to utilize certain finance firms was barred but General Motors was not forced to divest itself of GMAC. In 1941 and 1942 Chrysler unsuccessfully fought the bar against affiliation. On September 19, 1946, Ford appealed to have certain provisions of the consent decree lifted and in 1948 was successful. The Supreme Court reversed a decision of the lower court and lifted the ban on the Ford Motor Company's affiliation with a finance company.⁴

Financing of sales is also important in the sale of other transportation equipment. In the railway car and locomotive field, financing was particularly important during the depression when railroad credit was poor. General Motors as noted previously was able to enter the diesel locomotive field because it was willing to finance the sale of its product. ACF Industries, Incorporated, a producer of cars, has various financing affiliates. Fairbanks Morse has its captive finance company, the Municipal Acceptance Corporation. Pullman, Incorporated, also has a financing subsidiary to aid in the marketing of its railroad products.

The largest producer of truck trailers, Fruehauf Trailer Company, established a financing subsidiary in 1948, the Fruehauf Trailer Finance Company. In 1956 it established the Trailer Acceptance Company in Canada to finance the sale of its Canadian subsidiary's output. Trailmobile, a subsidiary of Pullman, Incorporated, established the Trailmobile Finance Company in 1955. Fruehauf and Trailmobile are estimated to control about 60 per cent of trailer production with the remainder of the market divided among many small producers. Since so many truck operators are relatively small and equipment comprises the major cost of entry into this industry, credit is extremely important. Fruehauf has offered its customers a seven year financing plan, roughly parallel in time to equipment life, and longer than regular financing agencies have offered. Smaller trailer companies have been unable to offer similar terms. On August

3. 121 F. 2d 376.

4. 335 U.S. 303. The court in lifting the ban argued that so long as the interdiction had not been decreed against General Motors the government must prove its case rather than rest on a consent decree. If the government wished to outlaw the Ford arrangements with a finance company, it had to establish its case in court. Thus, the reversal still left Ford with the possibility that if it did affiliate, the government could reopen its case and attempt to prove the illegality of the arrangement. The financing of Ford agricultural machinery output is described elsewhere.

17, 1956, the Federal Trade Commission charged the Fruehauf Trailer Company, among other things, with using its financing subsidiary to lessen competition and unfairly diverting trade to Fruehauf from its competitors.⁵ This case has not been settled as yet.

Truckers themselves, through the American Trucking Associations, Incorporated, are investigating a plan to establish an industry-sponsored credit pool to help truckers. Such a pool established by truckers and suppliers would not only help purchasers but would help the many small producers of truck trailers.

In addition to making passenger cars the big three auto producers are also major truck producers. General Motors' other financing subsidiary, Yellow Motors Acceptance Corporation, exists primarily to finance its truck sales. There are about eleven producers in the automobile industry which produce trucks but not passenger cars. Among these producers are included International Harvester which operates a finance affiliate. In 1955 Mack Truck arranged a revolving credit with forty-eight banks to finance the sale of its products. White Motor Company handles installment paper as ordinary receivables.

The aircraft firms have less need to finance the sale of their products since the government is the largest single purchaser, and air lines have in the past arranged their own financing. The two major producers of aircraft for private use have financing subsidiaries: Cessna organized a financing affiliate in 1955, and Beech Aircraft organized an affiliate in 1956. Lockheed has had an affiliate since 1943; Bell Aircraft's major stockholder is a finance company.

Sales financing is of particular importance in consumer durable goods and agricultural machinery. Manufacturers of consumer durable goods have been particularly active in the establishment of finance affiliates.

General Electric Corporation has had a financing subsidiary for many years. Its major expansion, however, has occurred since World War II. American Motors' Kelvinator Division has had a financing subsidiary to finance the sale of its appliances since 1934. The Westinghouse Credit Corporation was organized in 1956 with initial capital of \$10 million to help the company's appliance dealers obtain financing. Philco established a financing affiliate in 1954, with an initial capitalization of \$5 million and its line of credit is now \$25 million.

AVCO Manufacturing Company announced plans for a multi-

5. In the matter of Fruehauf Trailer Co.: Federal Trade Commission Docket No. 6608, Aug. 17, 1956.

million dollar financing plan for its dealers to meet the extended terms offered by competitive producers but no further information is available at this time. Borg Warner, producer of Norge appliances, established BW Acceptance Corporation in 1953. Various other manufacturers of heating equipment also have finance firms.

In the agricultural machinery field credit has always been of major importance and again we find the leading firms operating financing subsidiaries. Agricultural machinery producers in the past extended credit to their dealers and helped dealers to carry customer accounts. This is still the method used by Deere and Company, the Oliver Corporation, and the Minneapolis Moline Company. The largest firms in this industry, however, have organized sales finance firms. International Harvester, the largest producer of agricultural machinery, established its financing affiliate in 1949. Caterpillar Tractor established its affiliate in 1954; Cockshutt Farm Equipment, Limited, established the Farmers Financial Agency in 1955 and Allis-Chalmers established the Allis-Chalmers Credit Company in 1956. The J. I. Case Company has indicated that it, too, will organize a financing subsidiary.

Ford's activities in the agricultural machinery industry and the financing of such products is of interest in the light of Ford's experience in automobile financing. Various Ford Motor Company officials, though not members of the Ford family, including Ernest R. Breech, John S. Bugas, Lewis D. Crusoe, John R. Davis, William T. Gosset and Delmar S. Harder, own more than one-half of the stock of the Dearborn Motors Corporation which distributed Ford tractors and manufactured agricultural machinery. The company has, as a wholly-owned subsidiary, the Dearborn Motors Credit Corporation which finances tractors and other farm equipment at wholesale and retail. On July 31, 1953, Ford purchased the assets of Dearborn Motors Corporation, excluding Dearborn Motors Credit Corporation. The latter firm has continued its finance business and has financed more than half of the wholesale sales of Ford tractors and farm equipment and a smaller portion of the retail sales of such equipment. Thus, the Ford Motor Company does not own a finance company, but the Chairman of the Board of the Ford Motor Company and five vice-presidents own one which finances a goodly proportion of Ford agricultural implement output.

In the road machinery industry credit will undoubtedly be a major determinant among firms selling equipment for the new road building program. Contractors are generally small firms and in bidding for jobs must indicate financial soundness. Rental, long term

lease and extended time sales for equipment will be sought by this industry. Those firms with finance affiliates will be in a strong position to bid for this market. It should be noted that the large truck and agricultural machinery producers also build road machinery equipment, and also own finance companies, unlike smaller producers of specialized equipment.

In addition to manufacturers supplying credit, sufficiently large retailers have also entered the field of consumer credit. Various manufacturers who operate retail stores have been previously mentioned. In addition, the mail order houses have pushed time sales.

Sears, Roebuck recently found it advantageous to establish a sales finance company, Sears' second venture in this field. It first entered the installment field in 1911. When the Federal Housing Administration was established Sears organized the Sears Finance Company to provide FHA improvement financing. When banks entered this field, Sears retired, selling its accounts to banks and its financing affiliate became inactive at that time. It continued to finance installment sales, however, and the growth of such sales has been impressive. In the nine month period ending October 31, 1956, installment sales comprised 44 per cent of total sales.⁶ In 1937 Sears began selling some of its installment paper to banks. In 1952 as installment sales continued to rise, Sears borrowed \$200 million from various banks to finance the time payment accounts held rather than sold.⁷ On January 31, 1952 installment balances outstanding were \$493 million and by October 31, 1956, these balances had grown to \$973 million and the number of accounts from 5.7 million to 8.1 million. With the maturity of the bank loans Sears established the Sears Roebuck Acceptance Corporation, wholly owned by Sears. Its initial capital was \$35 million plus \$50 million in debentures. Sears is particularly well situated to enter the field of consumer credit. It has extensive local representation in the form of retail stores and catalogue offices.

Montgomery Ward finances its own credit accounts and at present holds in excess of \$200 million of such accounts.

SUMMARY

It is probable that monetary authorities have little interest in the competitive effects of captive finance companies and that anti-trust authorities have little interest in their monetary effects. Each problem is significant in its own right but together they add up to the general problem of size in our economy.

6. Prospectus, Sears Roebuck Acceptance Corporation, Jan. 10, 1952, p. 2.

7. *Moody's Industrial Manual*, 1955, p. 280.

If we are to experience a continuing expansion of captive finance firms, general credit controls will become increasingly inadequate to limit expansion of installment credit and the only control methods that could be effective in dealing with the situation created by the captive finance firms are direct credit controls. Furthermore, a critical study of the captive finance firm is needed to determine whether this is a more powerful tool to limit competition than has heretofore been utilized.

The captive finance firm is a growing influence in the credit structure of the country. The basis upon which these firms operate, their very reasons for existence, indicate that they are more effective from the parent firm's point of view when countering national monetary policy than when paralleling it. There are also reasons to believe that these firms are less responsive to general credit controls than are independent finance companies. Secondly, since the captive finance firm can be effective only if it has large sources of credit, its successful use is restricted to the large firm. As such it is another competitive advantage of size.

PAUL H. BANNER.

ST. LOUIS, MISSOURI

LOCATION AND THE THEORY OF PRODUCTION*

By LEON N. MOSES

This paper deals with the theory of the firm in a spatial setting. The paper is motivated by a belief that modern location theory still assumes a linear production function and in this respect has not progressed beyond Alfred Weber's original formulation of the location problem.¹ My objective is to place the theory of location within the main body of economic theory. More specifically, I wish to make the theory of location an integral part of the theory of production and to investigate the implications of factor substitution for the locational equilibrium of the firm. My main conclusion is that profit maximization requires a proper adjustment of output, input combi-

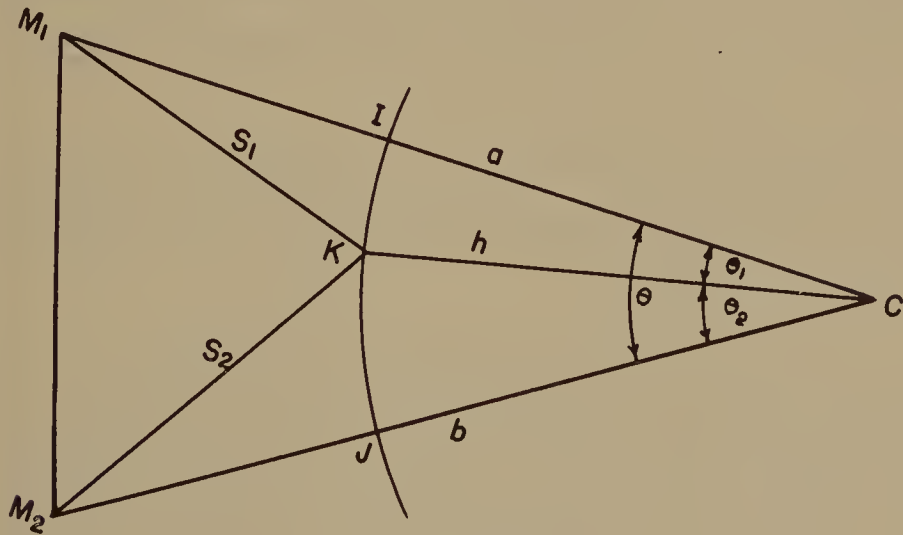


FIGURE I

nation, location, and price. Moreover, the optimizing values of these three variables can be determined with analytical tools derived directly from traditional economic theory. There is no need for much of the esoteric paraphernalia sometimes employed by location specialists.

* While assuming responsibility for any errors, the author wishes to express his gratitude to James M. Henderson, Marie McCarthy, Alan Strout, and Mary Sullivan for their helpful suggestions.

1. *Alfred Weber's Theory Of The Location Of Industries*, trans. C. J. Friedrich (The University of Chicago Press, 1929).

Most of the theory will be developed for the simple case of a firm which employs two transportable inputs to produce a single product that is sold in a single market point. Figure I depicts the locational problem. M_1 and M_2 are the sites of the two materials and C is the market point. The distances M_1M_2 , M_1C , and M_2C are known. The latter two distances will be referred to respectively as "a" and "b." The values of the included angles are also known, though we will be concerned only with the angle θ . Before enumerating the remainder of the notation it should be mentioned that the geometry is incidental to the concepts that emerge.

Notation:

- P_1 , the price of the first input at its source
- P_2 , the price of the second input at its source
- r_1 , the transport rate on the first input
- r_2 , the transport rate on the second input
- s_1 , the distance from M_1 to the locus of production of the final product
- s_2 , the distance from M_2 to the locus of production of the final product
- P'_1 , the price of the first input *delivered* to the locus of production of the final product
- P'_2 , the price of the second input *delivered* to the locus of production of the final product

I assume that inputs are sold f.o.b. Their delivered prices are, therefore, equal to price at the source plus full freight. If inputs are sold according to some other geographical pricing technique, delivered prices may be greater or less than f.o.b. prices, but this will not invalidate our analysis.

At the outset we deal with a partial equilibrium problem in which the distance that the final product must be shipped is held constant. Thus, an arc is described in Figure I from C which cuts the triangle at I and J . This arc is a segment of the circle with center at C and radius h . All points along the arc are a fixed distance from C and — initially — are the only ones which are considered as possible locations for the plant. With transportation cost on the final product fixed, we can concentrate on the problem of factor substitution as affected by variations in transportation expenditure. We do this by focusing attention on the ratio of delivered prices along the arc. In other words, each small move along the arc from I towards J increases the distance which the first raw material must be shipped, decreases the distance which the second raw material must be shipped, and alters the ratio of delivered prices.

In order to achieve a more rigorous statement of the problem, the angle θ is treated as being composed of two angles, θ_1 , and θ_2 .

θ_1 is permitted to take on all values between zero and θ . θ_1 , being assigned some value, θ_2 is equal to θ minus θ_1 . It is the variation in the angle θ_1 which defines the moving point that is the locus of production. For example, if θ_1 is equal to θ , then J is being considered as the production site. Suppose a value is assigned to the angle θ_1 which defines the point K on the arc. Join the points M_1 and M_2 with K . By the law of cosines, the distances from the two material sites to K can be determined:

$$(1) \quad s_1 = \sqrt{a^2 + h^2 - 2ah \cos \theta_1}$$

$$(2) \quad s_2 = \sqrt{b^2 + h^2 - 2bh \cos (\theta - \theta_1)}$$

The letters a , b , and h in the above equations represent the known distances M_1C , M_2C , and KC , respectively.

Since base prices and transportation rates are known, and since we have assumed that the materials are sold on an f.o.b. basis, the prices of the two materials delivered to K can also be determined.

$$(3) \quad \begin{aligned} P'_1 &= P_1 + r_1 \sqrt{a^2 + h^2 - 2ah \cos \theta_1} \\ &= P_1 + r_1 s_1 \end{aligned}$$

$$(4) \quad \begin{aligned} P'_2 &= P_2 + r_2 \sqrt{b^2 + h^2 - 2bh \cos (\theta - \theta_1)} \\ &= P_2 + r_2 s_2 \end{aligned}$$

We may also form a ratio between equations three and four:

$$\frac{P'_1}{P'_2} = \frac{P_1 + r_1 s_1}{P_2 + r_2 s_2}$$

What is this ratio? *It is the constant slope of the system of iso-outlay lines when production takes place at K .* There is no ambiguity concerning these iso-outlay lines. Each point along the arc IJ is characterized by a definite ratio of *delivered* prices on the two inputs. This ratio defines the slope of the system of iso-outlay lines at that point.

Two iso-outlay lines are shown in Figure II. Both represent the same total expenditure — taking into account base prices, and transportation — on the two material inputs. However, AB is one of the system of iso-outlay lines associated with production at I whereas DE is one of the system of iso-outlay lines associated with production at J . In the figure, emphasis is given to those portions of the two lines which make up the line AFE . The reason for the emphasis is that up to the point F , outlay line AB (and hence, location at I) offers the possibility of purchasing the same amount of one

input, say M_1 , as outlay line DE (representing location at J) but more of the other input. For example, consider the points V and Z , both involving the same expenditure. V lies on the iso-outlay line corresponding to location at I while Z lies on the iso-outlay line corresponding to location at J . The figure indicates that if the firm locates at I it will be able to purchase the same quantity of M_2 as it could if it located at J , but that it can acquire a greater amount of M_1 (equal to VZ) with total expenditure the same. To the right of point F the

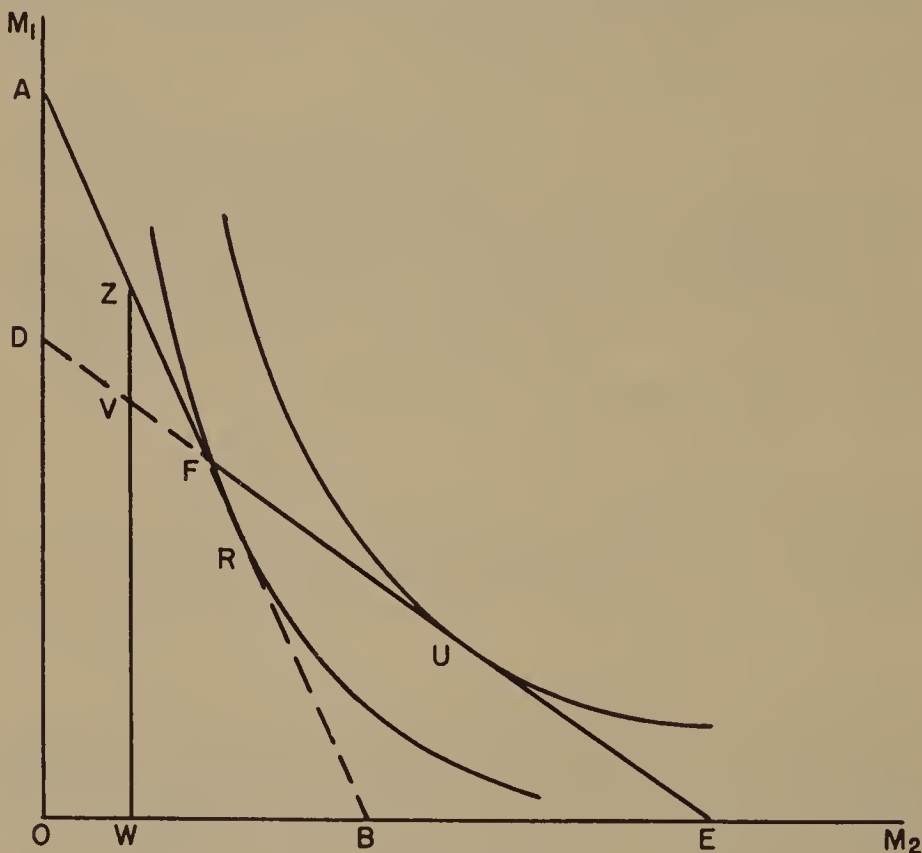


FIGURE II

situation is reversed. Location at J permits the same quantity of M_1 to be purchased as at I but more of M_2 , expenditure given. In any particular case, the isoquants determine the relative profitability of locating at J or at I .

Suppose the firm decides to expend upon production an amount which corresponds to the iso-outlay lines of Figure II. In this case — questions of demand elasticity aside — the profits of the firm would be greater if it located at J rather than at I since the same total expenditure yields a higher output at the former. Two points of

tangency, R and U , are shown in the figure. R is a tangency point produced by an isoquant and that portion of the outlay line AB which lies below F . Clearly, the firm is better off if it locates at J and uses the combination of inputs associated with the tangency point U since the latter lies on a higher isoquant than R .

Up to now only two locations along the arc IJ have been considered, and therefore, only two differently sloped iso-outlay lines appear in Figure II. If each of the infinite number of points along the arc IJ is treated as a possible location, then the discontinuous line AFE becomes a smooth curve. The reason for this is that each point along the arc has a unique ratio of delivered prices. We will call this smooth curve the locational iso-outlay curve. *Each point along this curve corresponds to a particular location along the arc IJ , and shows the one combination of factors which the firm would purchase at that particular location, given the dollar expenditure, transport rates, and base prices of the inputs.* It must be emphasized that if continuous spatial substitution is possible (i.e., all points along the arc IJ may become the site of production), then there is only one combination of inputs which is optimal for each location. All other combinations are of the same sort as those depicted by R and V in Figure II. In other words, the locational iso-outlay curve represents the results of an optimization process. Uneconomic combinations of inputs do not appear in it. Which combinations of inputs are uneconomic for each location can be decided without reference to the production function. The relevant factors are distance relations, the spatial structure of transport costs, and the prices of inputs at their sources.

The nature of the locational iso-outlay curve can be brought out still more clearly by analogy with the familiar case of production with variable input prices. Consider a producer, who, by varying the manner in which he apportions a given expenditure on two inputs, can reduce the price of one but only at the expense of an increase in the price of the other. The firm seeking a location along the arc IJ is in a similar position. It can reduce the *delivered* price of M_1 by selecting a location along the arc closer to the source of this input, but only at the expense of an increase in the delivered price of M_2 . As an aside we may also mention that if each point along the arc IJ has its own ratio of delivered prices, then each point also has its own total cost curve. All of these cost curves together form what we may call the "locational planning curve," to draw another analogy from traditional theory.

Figure II considers only one level of expenditure. If different levels of expenditure are considered, then a system of locational

iso-outlay curves will be generated. The dashed curves of Figure III depict such a system. All of these curves pertain to locations along the arc IJ of Figure I, but each represents a different level of expenditure upon the inputs. The solid curves of the diagram are isoquants. The slope of the line KL defines a point along the arc IJ of Figure I.

Suppose the firm wished to deliver a fixed number of units of the product to the market, for example 1000. As in traditional theory, optimality is characterized by the tangency condition, i.e., at B in

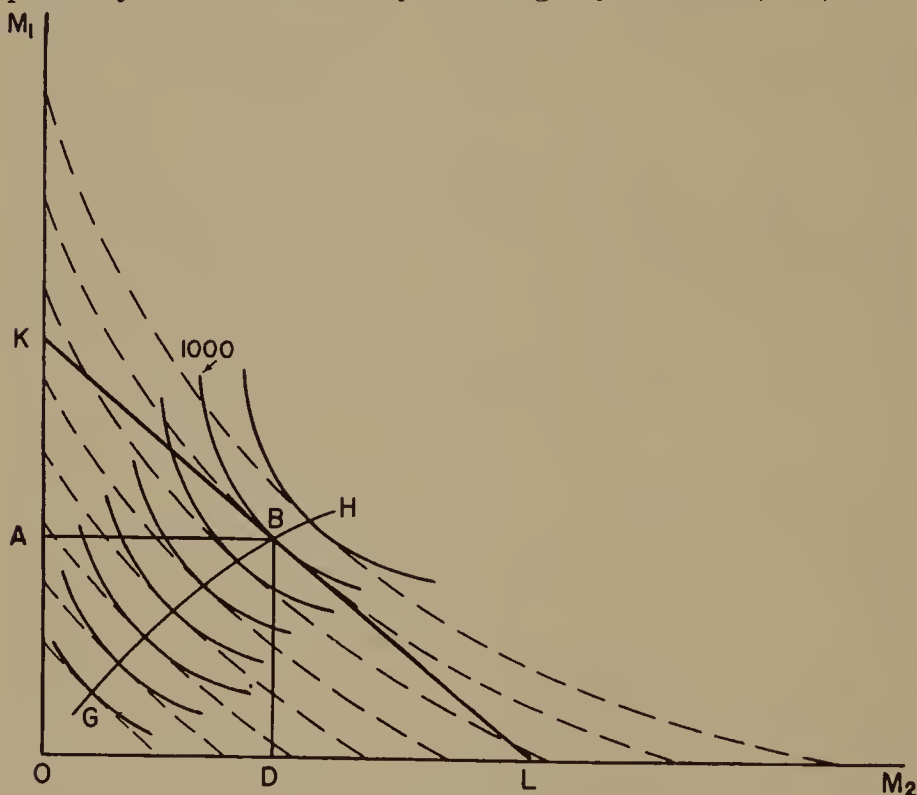


FIGURE III

Figure III. However, the point B represents not only a particular combination of inputs — as in traditional theory — but also a particular location along the arc IJ . The firm chooses that location at which the ratio of delivered prices is equal to the slope of the 1000 unit isoquant. In this location the firm will use OA units of M_1 and OD units of M_2 . The marginal rate of outlay substitution will be equal to the marginal rate of input substitution. In other words, within the dimensions of the problem as so far formulated, the optimum location along the arc IJ of Figure I is the point at which total expenditure is a minimum. At this location the factors will be combined in such a way that the

ratio of their marginal productivities is equal to the ratio of their delivered prices. It would be purely by chance if total transportation costs were a minimum at this point. This conclusion anticipates the brief survey of location literature found in the final pages of this paper.

In analyzing the location problem of an individual firm, Weber, Hoover, and Isard begin by assuming that there are no geographic variations in the prices or qualities of inputs at alternative sources.² From this assumption they derive two conclusions: (1) that there are no geographic differentials in production costs, and hence; that (2) the optimum location is the point of minimum transportation costs. However, the two conclusions do not follow from the aforementioned assumption alone. They also involve an additional one, namely, that the production function is linear. There are no alternative sources for inputs in the example with which we have been working. However, we have permitted factor substitution and therefore have altered the conditions for optimality in location. This difference is not the only one which emerges.

A number of points of tangency between locational iso-outlay lines and isoquants is shown in Figure III. These points trace an expansion path, labelled *GBH*. It differs from the expansion path of traditional theory in that each point along its length refers to a location (along the arc *IJ* in the present case) as well as a level of output and a combination of inputs. This expansion path brings an interesting point to light, and one which, to the author's knowledge, is not made clear in previous works on the theory of location. *If the production function is not homogeneous of the first degree there is no single optimum location along the arc IJ. The optimum location varies with the level of output.* Moreover, if continuous spatial, as well as factor, substitution is possible there will be a different optimum for every level of output. This being so, demand considerations should be introduced into the analysis. However, before doing so, there are two additional points that require consideration.

First, as we have drawn the iso-outlay lines and isoquants in Figure III, there is a single point of tangency for each output or each level of expenditure. There is no a priori reason for this condition. In actuality, there may be more than one point of tangency, i.e., alternative locations may exist which are equally good for producing a given output or expending a given amount upon production.

Second, we have drawn the curves of Figure III so that the tangency points are equilibrium points. There does not appear to be

2. Specific footnote references to the works of these authors appear below.

any reason why this must be so. The isoquants can be tangent to the outlay lines from below. Whether they are or not will depend in any particular case on the production function, base prices of materials, transportation rates on materials, and their geographic position relative to one another. We are still considering the partial problem of location along the arc IJ of Figure I. In this case, if the isoquants are tangent to the outlay lines from below, the firm would locate either at I or J . In order to bring out a somewhat different point, let us postulate that transportation costs on the final product are negligible, and that the firm is free to locate anywhere in the locational triangle of Figure I. In this case, if the inputs are substitutable in the extreme — by which we mean that the product can be produced entirely with one or the other of them — then the firm will locate at the source of one of the inputs and use none of the other. If the inputs are not substitutable in the extreme, the firm will be drawn close to the source of one of them, and use relatively more of it than the other.

Above, it was stated that demand considerations should be introduced into the analysis. In order to save space this issue will be treated in conjunction with a number of other factors so far ignored. First, regional variations in the money price of a given input will be introduced. Second, the effect of transportation cost incurred in shipping the product to the market will be taken into account. As to the latter, the reader will recall that up to this time we have been wholly concerned with substitution between inputs, transportation on the product being held constant. Now we wish to allow this latter distance to vary. In effect, we will analyze a problem which is equivalent to allowing the arc of Figure I to shift either closer or farther away from the point of consumption.

For these purposes a still more simplified locational figure will be employed, the straight line.³ We assume a productive process that requires two inputs, a single localized raw material, M , and labor. Labor is available both at the site of the raw material and at the market for the product, C . However, its money price is lower at the latter. The quality of the labor at the two places is identical. The price differentials implicit in the problem give rise to the V-shaped iso-outlay lines of Figure IV. The more steeply sloped portions of these lines represent combinations of inputs which can be purchased for a given expenditure if the firm locates at M . The more gently sloped portion of each outlay line shows the combination of inputs which can be purchased for the same expenditure if the firm locates at

3. Since the locational figure is so simple there is no need for a figure.

C. The difference in slope arises from the fact that at *M* the price of the raw material is lower than at *C*, while the price of labor is higher than at *C*. Thus, for a given expenditure, more of the material and less labor can be acquired if the firm locates at *M*. The reverse is true at *C*.

A system of isoquants and the expansion line formed by their points of tangency with the locational iso-outlay lines are shown in Figure IV. Each point of tangency⁴ indicates two things: the best



FIGURE IV

location for producing a given output;⁵ the best combination of inputs for producing that output at that location. According to the expansion line, *GBHT*, location at the raw material site, *M*, is a superior location for outputs up to 800 units, whereas location at the

4. In order to keep the diagram simple only one equilibrium location and combination of inputs is shown for each output. There is no reason why it is not possible to have a condition of double tangency for some outputs. This condition would mean that the two locations are equally good for producing the outputs in question. The cost of distributing the product would then become the crucial factor in determining the optimum location.

5. Optimality in location is indicated by the portion of a given locational iso-outlay line where tangency with an isoquant occurs.

market site, C , is superior for outputs of 1000 or more units. This result is brought out more clearly in Figure V, which contains two total cost curves.⁶ TC_m is the total cost curve the firm would have if it located at the site of the raw material. TC_c is the total cost curve the firm would have if it located at the market. The figure also depicts two total revenue curves. These curves will help us to appraise the effect of transportation cost on the product.

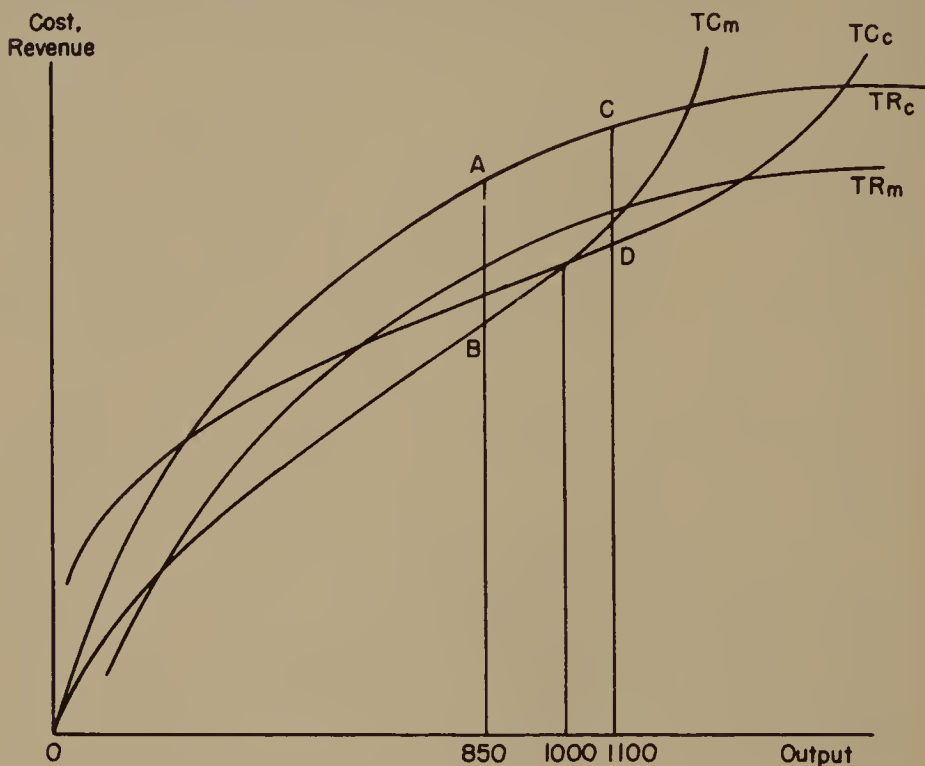


FIGURE V

TR_c is the total revenue curve for a monopolistically competitive firm as it exists in the market. (We do not present it, but the argument can readily be formulated for the purely competitive case.) TR_m is the same total revenue curve reduced by the amount of transportation cost the firm would incur on each output in shipping this output to the market. In other words, if the firm locates at the site of the raw material, it will not realize the price the product sells for in the market but rather this price less the per unit transportation

6. These curves are two of the family of such cost curves (comprising the locational planning curve) that we would have to consider if all points between the raw material site and the consumption point were treated as possible locations.

charge. The curve TR_m is derived by multiplying the prices net of transportation costs by the corresponding outputs.⁷

From the figure, we see that if transportation cost on the product were zero (and TR_m therefore identical to TR_c), the firm would maximize its profits (AB) by locating at the site of the raw material and producing 850 units. However, as we have drawn the total revenue curves, transportation cost on the final product is not zero. The firm maximizes its profits (CD) by locating at the market and producing 1100 units. It should be understood that a shift in the demand relation from which the total revenue curves are derived could alter both the optimum output and location. Thus, the optimum location is seen finally to depend on the following factors: base prices on inputs; transportation rates on inputs and on the final product; the geographic position of materials and markets; the production function; the demand function. We consider these the primary factors, but they do not comprise a complete list. For example, selling costs can influence the demand function and therefore influence location.

As stated earlier, the main objective of this paper is to make the theory of location an integral part of the theory of the firm. Since this is not the first such effort, I will now briefly indicate the similarities and differences between the present paper and several others.

In 1928 Andreas Predöhl commented that the problems of location had been barely touched upon in systematic theories of economics and that these theories must be applicable to the problem of location: "... they must contain the undeveloped principles of location."⁸ In his view, the problem of location was the problem of combining the means of production by different technical methods and therefore should be amenable to analysis by the principles of substitution.

There is little in Predöhl's paper with which the present author disagrees. Rather, my criticism is that he contented himself with a bare statement of the problem. He failed to develop and explain the mechanism through which factor substitution works when transportation costs are introduced into the theory of production; nor did he develop the substitution relation between transport expenditures for gathering raw materials and those for shipping the product to the market. Nevertheless, the present paper may be viewed as an effort

7. One could just as easily proceed by thinking of price as being set at the place of production and then reducing the quantity demanded on the basis of transportation charges added to this price.

8. Andreas Predöhl, "The Theory of Location in its Relation to General Economics," *The Journal of Political Economy*, XXXVI (1928), 371-90.

to make explicit, propositions which are implicit in Predöhl's paper and to investigate their implications. Walter Isard's recent volume may be similarly viewed.⁹

Professor Isard works out some elaborate substitution problems in his study. However, I do not believe he has achieved a generalization of location theory. Rather, he has presented us with a novel and rigorous formulation of Weber's theory of transport and other orientation, which is location under the assumption of a linear production function. In one place, Isard almost states this as his objective. He claims that the concept of a transport input — defined as the movement of a unit weight over a unit distance — “. . . will enable us to fuse much of traditional Weberian location doctrine and modern production theory.”¹ Weberian location doctrine is concerned almost entirely with the case of constant coefficients of production whereas the theory of production focuses interest on cases in which inputs are substitutable.

The term “substitution” occurs frequently in Isard's volume but the concept is employed in the context of one, or a combination, of the following situations:

- (1) substitution between transport expenditures on various mobile inputs, the weights of each of them required per unit of output held constant;
- (2) substitution between transport outlays for mobile inputs and outlays for some immobile input such as labor, the weights of the mobile inputs as well as the quantity of labor required per unit of output held constant;
- (3) substitution between transport outlays for gathering mobile inputs and outlays for immobile inputs on the one hand, and transport outlays for shipping a unit of the product to market on the other, with the weights or quantities of the inputs required per unit of product held constant.

Isard's adherence to the assumption of a linear production function leads him to the conclusion that, for the case in which there are no alternative sources of inputs or no regional variations in their base prices, there is a single best location and this occurs where transportation costs on the inputs — those on the product held constant — are a minimum. I have indicated that this result is a special case. If inputs are substitutable, there is no single optimum location. The optimum location then depends on the scale of operations and other things mentioned earlier.

Isard is, of course, not unaware of the importance of scale economies. However, they are brought explicitly into the analysis

9. Walter Isard, *Location and Space Economy* (New York: jointly by The Technology Press of Massachusetts Institute of Technology and John Wiley and Sons, 1956).

1. *Ibid.*, p. 91.

only after a location has been decided upon. Thus, in discussing market and supply areas, he employs Professor Hoover's concept of a margin line. This line shows how delivered price at the edge of a market varies as the extent of the market, and hence output, varies, the geographic position of the firm being fixed.

Much the same approach is found in Hoover's first volume on location.² Constant costs are relied upon in determining the optimum location. The usual U-shaped cost curve is brought into the analysis only after the firm has been located. However, Hoover appears to be disturbed by an apparent conflict between reality and theoretical conclusions based on Weberian location doctrine. Reality presents numerous instances in which firms in a given industry have different orientations. Weberian location doctrine favors the conclusion that there is a single best location for each type of activity. In order to bring the theory into closer accord with reality, Hoover introduces such things as variations in the quality of a given input at alternative sources, differences in transport rate gradients in different directions, etc. He uses these differences to derive zones of orientation.³ Some of these zones are characterized by orientation to market, some by orientation to one or another raw material. Variations in the quality of inputs and peculiarities in transport rate structures are undoubtedly important in explaining realistic location patterns. However, it is equally important to introduce factor substitution. Hoover does recognize this in his later volume.⁴

The theory of the location of economic activity studies individual economic units, as well as groups of interrelated units and regional economies. The present paper has been concerned with the former. Its object has been to make the theory of location, as applied to the individual firm, an integral part of the general theory of production. Towards this end, we have given emphasis to the inseparability of three problems: the optimum output; the optimum combination of inputs; the optimum location. The approach developed in the present paper facilitates analysis of the spatial aspects of many problems traditionally considered in theory: i.e., the implications for location of variable input prices, of discrimination between markets, etc. The idea of a break in the expansion line — developed in Figure IV — may shed light on the establishment of branch plants.

Finally, the paper contains an implicit warning against a narrow technological approach to location, such as is practised by some of

2. Edgar M. Hoover, Jr., *Location Theory and the Shoe and Leather Industries*.

3. *Ibid.*, pp. 48-51.

4. Edgar M. Hoover, Jr., *The Location of Economic Activity*, pp. 44-46.

the largest engineering consulting firms. This engineering approach tends to restrict, to an unwarranted extent, the range of possible types of locations which individual plants may consider. It may also delimit too severely the range of possible types of industries which underdeveloped areas are advised to consider in their development plans. We are not suggesting that it is feasible or even necessary to work with continuous factor substitution in solving practical location problems. However, it does seem that it would at least be worthwhile to investigate the possibility of producing according to alternative linear production functions.

LEON N. MOSES.

HARVARD UNIVERSITY

THE EFFECT OF DEVALUATION ON THE PRICE LEVEL*

By EGON SOHMEN

I. Argument that depreciation of an overvalued currency must entail a rise in the domestic price level, 273. — II. Examination of this argument, 277. — III. Differences between various restrictive policies, 281.

I

The present paper deals with the widespread belief that depreciation of an overvalued currency must necessarily entail a rise in the domestic price level of the depreciating country, unless it is accompanied by deflationary fiscal and monetary policies. It is no exaggeration to say that this belief has served as perhaps the single most potent argument against an adjustment of overvalued currencies. It is, of course, a stock-in-trade of the convinced advocates of exchange controls;¹ however, it has usually also been accepted by outspoken foes of all restraints on free currency convertibility.²

In a passage of his League of Nations monograph, Ragnar Nurkse recognizes the validity of the argument advanced here,³ in contradic-

* The major argument advanced here had its origin in conversations with Professor Charles P. Kindleberger. I have greatly benefited from a number of suggestions by Professor Gottfried Haberler which are gratefully acknowledged.

1. "[W]hen world prices are rising, or even when they are more or less steady, . . . a lowering of exchange rates would cause internal prices to rise absolutely, and not merely relatively to prices in the outside world. The rise would not be confined to the prices of imported goods, but would spread gradually throughout the country's economy. . . . In any other environment than that of world deflation, a substantial exchange depreciation must be a strong price-raising influence." Sir Hubert Henderson, "The Function of Exchange Rates," *Oxford Economic Papers*, I (1949), 7.

2. "Devaluation raises the domestic prices of imports and of exports; and it will tend to raise the prices of import substitutes, of potential exports, and of intermediate goods required for their production. Thus, unless the monetary authorities restrict credits in order to force price reductions in other sectors of the economy, the price level will be somewhat increased as a result of the devaluation." F. Machlup, "Relative Prices and Aggregate Spending in the Analysis of Devaluation," *American Economic Review*, XLV (1955), 262.

3. "More specific objections to devaluation were based on the fear of a rise . . . in import prices . . . in terms of domestic money. The fact was sometimes overlooked that a rise in the prices of imported goods is bound to occur when exchange restrictions create an artificial scarcity of such goods on the home market, unless the goods are subjected to price control and rationing." *International Currency Experience*, League of Nations, 1944, p. 167.

tion to the opinion he voices in an earlier chapter of the same book.⁴ To my knowledge, the only theorist who has recently argued that the traditional belief is not necessarily correct is Dr. Edward M. Bernstein, Research Director of the International Monetary Fund.⁵ The opinion prevalent among theorists is shared by the majority of the practitioners of economic policy, and, of course, by the layman public.⁶

On superficial examination, the case appears deceptively straightforward. A depreciated exchange rate, it is held, implies a price increase for all those imported goods whose price in foreign currency does not simultaneously fall in the same proportion. The inflationary

4. "[T]he fall in the exchange rate became literally the leading factor in the mechanism of inflation, driving up the cost of living and creating an irresistible pressure for wage adjustments, which in turn called forth demands for additional currency on the part of the government as well as business men." *Op. cit.*, p. 115. The chapter deals with "Exchange Depreciation in the Early 'Twenties."

In correspondence, Professor Nurkse quite correctly points out to me that the latter statement concerns the effects of capital flight induced by an inflationary trend, and therefore does not necessarily conflict with the first quotation. The real point to be emphasized here is that artificial pegging of exchange rates can under such circumstances never prevent a speculative capital outflow, but will, on the contrary, further encourage it. Capital flight can be effectively prevented only by exchange controls, and the latter are equally feasible (although less frequently practiced) when the currency is allowed to depreciate. Nurkse's work provides an argument for monetary stability, and for controls over capital movements when inflation is unavoidable, but cannot be used to justify the continuation of an artificial exchange peg under inflationary conditions, as is frequently maintained.

5. E. M. Bernstein, "Strategic Factors in Balance of Payments Adjustment," *International Monetary Fund Staff Papers*, V (1956), esp. pp. 166-68. Dr. Bernstein was kind enough to make a letter available to me, written by him to Professor Haberler in March 1949, in which the reasoning of his later article is already clearly enunciated. As an empirical example for his main thesis, he points out that the French devaluation of 1948 (by 80 per cent) had a deflationary effect.

6. After the recent French devaluation, the *New York Times* reported from Paris: "The French government took action on several fronts today to prevent a general price rise following Saturday's virtual devaluation of the franc. . . . The first two days after devaluation have produced no significant price rises on purely domestic products. This was interpreted as evidence of the success of the official effort to minimize the importance of the monetary adjustment and, above all, to avoid use of the word 'devaluation'." (My italics) *New York Times*, Aug. 14, 1957, p. 9. The last sentence of the quotation illustrates another well-known fact: layman opinion in many European countries often sees not merely a more or less loose relationship between depreciation and a rise in the domestic price level, but occasionally tends to identify the word "devaluation" with an increase of prices in the same proportion. We may again quote from Nurkse's monograph: "Memories of that experience [the hyper-inflation in Germany and elsewhere in Europe during the early 'twenties] played no doubt an important part in the popular identification of devaluation with inflation in Central and Eastern Europe in the 'thirties. The German trade unions in 1931 threatened, in the event of devaluation, to demand a scale of money wages based on foreign exchange rates." League of Nations, *op. cit.*, p. 167.

effect is reinforced by the drain on domestic resources due to the induced increase of the country's exports. Only a deflation of the national income to match the reduction in aggregate supply on the domestic market, so the reasoning continues, can prevent the increase in the over-all price level.

The argument would be in accord with reality if commodity movements were always perfectly free and unimpeded at any arbitrarily fixed exchange rate, with accommodating capital transfers⁷ filling the gap in the balance on current account. The domestic price of imports would then indeed always be equal to their foreign price, multiplied by the exchange rate.

However, such a picture conflicts radically with the observed facts. No country whose currency is overvalued can ever count on an unlimited source of accommodating finance. As a consequence, it has to rely on tariffs, import quotas, or exchange controls to check the deficit in the balance of payments that would otherwise develop.

Whenever such interference occurs, however, external and domestic prices must diverge (compared on the basis of the prevailing foreign exchange rate). What the combined effect of these various influences on the average price level will be can only be guesswork without resort to more refined tools of economic analysis.

Upon reflection, it is evident that only the tools of welfare economics can provide a determinate answer. The following relatively obvious theorem will be used: If national income in money terms remains constant, the aggregate price level (measured in terms of any one of the known varieties of price index numbers with the exception of Laspeyres' index which *may* remain constant) must fall after an unambiguous improvement in potential welfare.

The proof is set out in a few lines of algebra. Subscripts 1 denote prices and quantities in the initial (inferior) situation, subscripts 2 refer to the situation after the change.

An unambiguous improvement in potential welfare implies both

$$(1) \quad \begin{aligned} \Sigma P_2 Q_2 &\geq \Sigma P_2 Q_1 && \text{and} \\ \Sigma P_1 Q_2 &> \Sigma P_1 Q_1 \end{aligned}^8$$

7. Throughout this paper, the term "capital transfers" will be used for all transfers which serve to finance an imbalance in a country's current account (goods and invisibles). They will therefore be meant to include all long- and short-term loans extended by private individuals, businesses, or the government; shipments or earmarking of gold, as well as purchases and sales of foreign exchange by the country's central bank or exchange stabilization fund. This rather wide definition of the term is adopted for convenience only.

8. This is equivalent to Scitovsky's double welfare criterion, first stated in his two articles in the *Review of Economic Studies*, IX (1941-42), in criticism of the earlier, incomplete Kaldor-Hicks criterion. For a detailed exposition of its

If we assume, as a first approximation, that the country follows a "neutral"⁹ monetary and fiscal policy in the sense that money national income is held constant, i.e., if

$$(2) \quad Y_1 = \Sigma P_1 Q_1 = Y_2 = \Sigma P_2 Q_2,$$

we obtain, after dividing the first inequality by $\Sigma P_1 Q_1$, the second one into $\Sigma P_2 Q_2$, and using (2),

$$(3) \quad \frac{\Sigma P_2 Q_1}{\Sigma P_1 Q_1} \leq 1 \quad \text{and} \quad \frac{\Sigma P_2 Q_2}{\Sigma P_1 Q_2} < 1.$$

The first expression in (3) is Laspeyres', the second Paasche's index. Both show a lower level of prices in the second (improved) situation. The equality sign in Laspeyres' index holds in the context of international trade when factors of production are perfectly immobile so that the pattern of production is not changed at all after any given modification of the pattern of foreign trade. Geometrically, the production possibility hypersurface has a kink at the point of actual production. However, this theoretical possibility in no way prevents the accrual of gains from trade, as Haberler has stressed.¹

It should be emphasized that I assume full employment and perfectly competitive conditions within each economy throughout this paper. In other words, it is assumed that the economy operates on its production possibility hypersurface at all times. If an economy is underemployed before its currency is depreciated, the demonstration that the price level need not rise is trivial. The same holds true if the Paretian optimum conditions are not fulfilled.

Parenthetically, it might be noted that the price index numbers calculated in practice concern not the aggregate of all commodities produced in an economy, but only a selected sample of them. The price index for such a sample and the one for the aggregate which meaning, see I. M. D. Little, *A Critique of Welfare Economics* (2d ed.; Oxford: Clarendon Press, 1957); or Paul A. Samuelson, "Evaluation of Real National Income," *Oxford Economic Papers*, II (1950). The improvement is "unambiguous" if both conditions are fulfilled simultaneously; but we can nevertheless call it a "potential" improvement only as long as we do not know whether the change is accompanied by a suitable redistribution of incomes to guarantee an actual improvement in accordance with the interpersonal judgments the society — or its dictator, or ruling body — has chosen to adopt.

9. This definition of a "neutral" policy, adopted here for lack of a more appropriate term, differs substantially from the terminology of other authors. See esp. J. E. Meade, *The Balance of Payments* (London and New York: Oxford University Press, 1951), pp. 47-49.

1. G. Haberler, "Some Problems in the Pure Theory of International Trade," *Economic Journal*, LX (1950), 228-29.

underlies our analysis may occasionally go in opposite directions. But it seems more plausible to take the aggregate index when we deal with movements of the price level in a purely abstract way.

II

The preceding spadework reveals clearly that depreciation can lead to a rise in the price level (defined in the above sense) if and only if it entails a decline in potential welfare according to the Hicks-Kaldor-Scitovsky welfare criteria. Seen in this light, the alleged necessity of depreciation having an inflationary effect appears already decidedly less plausible. In particular, we are alerted against the fallacy of using an administratively fixed exchange rate to compute the domestic price of imports. The really relevant factors for the welfare position of the country are the measures through which equilibrium in the balance of payments is brought about, once the exchange rate is (as assumed) held at a disequilibrium position. If the currency is overvalued, less can be exported than would otherwise be possible. Whether exchange controls, import quotas or the old-fashioned tariff are used to plug the resulting leak in the country's reserves, all have as their objective an artificial restraint on the country's imports.

The net result is that the pattern of production moves towards lines of comparative disadvantage. From this fact alone, the earlier classicists would have jumped to the conclusion that the country's welfare position must have deteriorated. Our corollary would then necessarily follow that depreciation to the equilibrium rate of exchange, combined with a removal of trade and exchange controls, must *lower* the domestic price level.

However, this view leaves out of account the terms-of-trade argument for tariffs. The unilateral imposition of trade restrictions, whether tariffs, import quotas, or exchange controls, will result in an improvement in the country's terms of trade, i.e., of the price ratio between its exports and imports in the *outside* world.² It has to be noted, first of all, that the corresponding *domestic* price ratio moves, of course, in the *opposite* direction.

Nevertheless, the terms-of-trade argument shows that the imposition of a sufficiently small trade restriction will bring about a rise in potential welfare, defined according to the criteria outlined above. Once the optimum tariff (or the "optimum quota") has been passed, however, potential welfare will decline again while the external terms

2. Provided the foreign country's reciprocal demand is less than infinitely elastic.

of trade continue to improve. As the country moves further toward its autarky position, a point is finally reached after which the level of potential welfare falls below the one attainable under free trade.³

The movement in the aggregate price level is, as shown above, directly linked to the changes in potential welfare: starting from free trade, a sufficiently small restriction will indeed lower the price level. However, the outcome is reversed once the level of the "optimum"⁴ degree of restriction is exceeded.⁵

Turning to the real world, what is the chance that the typical

3. The history of the terms-of-trade argument offers a bewildering pattern of discoveries and rediscoveries in which it is not always clear who invented what. My own research furnished the result that Auspitz and Lieben have to be credited with the earliest rigorous statement of the theorem (*Untersuchungen über die Theorie des Preises* (Leipzig: Duncker and Humblot, 1889), pp. 415-18 and Figures 73 and 74). The argument is carried surprisingly far; not only do they demonstrate the condition for the optimum tariff, but they are also aware of the symmetry between import and export duties (p. 418). Apart from Viner (*Studies in the Theory of International Trade*, pp. 592-93), their contribution is never mentioned in the literature of today, although it was acknowledged by Edgeworth ("The Theory of International Values," *Economic Journal*, IV (1894), 636). The only flaw a modern theorist would find in their argument is the fact that it is stated in terms of consumers' surplus, in accord with the practices of their day.

For later rediscoveries and refinements, see T. de Scitovsky, "A Reconsideration of the Theory of Tariffs," *Review of Economic Studies*, IX (1941-42), reprinted in the American Economic Association's *Readings in the Theory of International Trade*. A geometric demonstration that avoids the use of community indifference curves is given in Peter B. Kenen's recent article, "On the Geometry of Welfare Economics," this *Journal*, LXXI (Aug. 1957), 442-45.

Among papers in which the terms-of-trade effect of import restrictions in general (whether tariffs, quotas, or exchange controls) is analyzed, Sidney S. Alexander's "Devaluation Versus Import Restriction as an Instrument for Improving Foreign Balance," *International Monetary Fund Staff Papers*, Vol. I (1950-51), and J. M. Fleming's "On Making the Best of Balance of Payments Restrictions on Imports," *Economic Journal*, LXI (1951), 48-71, are best known.

4. The inherent ambiguity of the "optimum" tariff or quota should be noted at this point; it will in general differ for different patterns of income distribution.

5. To be quite precise, the Laspeyres index will show a reversal even *before* the optimum tariff or quota is attained. This is shown for the two-commodity case in Figure I. Under free trade, production is at P_1 , consumption at 1. A trade restriction of a certain level may result in production at P_2 and consumption at 2. The index number comparisons yield

$$(4) \quad \Sigma P_2 Q_2 < \Sigma P_2 Q_1 \quad \text{and} \quad \Sigma P_1 Q_2 > \Sigma P_1 Q_1.$$

Proceeding as above, we obtain

$$(5) \quad \frac{\Sigma P_2 Q_1}{\Sigma P_1 Q_1} > 1, \quad \frac{\Sigma P_2 Q_2}{\Sigma P_1 Q_2} < 1.$$

Laspeyres' index thus shows a higher price level after the trade restriction, even though this situation may still be preferable to the free-trade position according

deficit country is short of its "optimum" degree of restriction? The question can be settled decisively only on the basis of a detailed empirical investigation; but there seem to be very convincing reasons to conclude that it is practically nil.

It is worth emphasizing, first of all, that the impact of trade restrictions is additive: the welfare level will depend on the effect of *all* existing tariffs, quotas, and exchange controls taken together.⁶ It seems reasonable to assume that tariffs alone already exceed the optimum level for most countries. We remember that the optimum degree of restriction decreases the higher is the elasticity of foreign reciprocal demand. It is a fairly safe assumption that, for any single country, the reciprocal demand of the rest of the world (depending, as it does, not only on the demand for its exports, but in addition on the supply elasticity of its imports) will be relatively high.

Any additional trade restriction, coming on top of existing tariffs, can then only aggravate the country's welfare position. A currency depreciation which permits their removal will under these circumstances by necessity result in a fall of domestic prices.

The fall in the price level, furthermore, must be all the more pronounced *the larger is the extent of the depreciation* necessary to establish equilibrium in the exchange market, contrary to the conclusion which uncautious intuition and the accepted doctrine on this matter would suggest. A large depreciation implies that the currency was substantially overvalued, and that the import restrictions must have been all the more severe.

An illuminating corollary is obtained by using the link between welfare and price level changes in reverse. Suppose for a moment

to a certain welfare function (or functions) and for given patterns of income distribution.

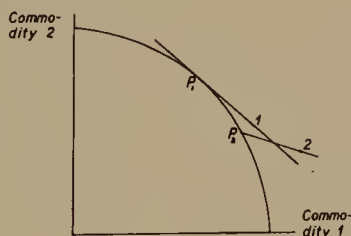


FIGURE I

6. I am indebted to Dr. Bernstein for this observation. The level of the optimum tariff is raised in the presence of foreign retaliation (Scitovsky, *op. cit.*, *Readings*, pp. 373-74). But the competitive game of raising retaliatory tariffs will always encounter a "point of no further return."

that, contrary to our conclusions, the traditional belief were true that a depreciation always had to result in a rise of the price level, and that it was all the more inflationary the more drastically a currency is devalued. It would then necessarily follow that a country's potential welfare (and hence conceivably the real income of every citizen after a suitable income redistribution) could always be raised by an administered appreciation of its currency, accompanied by whatever import restrictions are called for to keep the balance of payments in order. The rise in potential welfare would have to be the more pronounced the more overvalued the currency becomes. Pursuing the argument to its end, we deduce that a country's welfare potential must reach a maximum when the exchange rate reaches a level at which all exports are cut off (any additional devaluation does not cause any further change in the relevant variables). This *reductio ad absurdum* should amply illustrate the glaring defects of the accepted doctrine.

The associated welfare connotations give an indication of the losses in real incomes that the continuation of the present exchange rate maladjustments must inflict on all the countries of the world. It would also appear from our analysis that, again in sharp contrast to accepted beliefs, the countries whose currencies are the most overvalued are those who bear the brunt of the burden — and they are usually the ones who can least afford it.

That a depreciation of overvalued currencies — or, more precisely, the removal of barriers to trade it permits — is likely to have a deflationary effect does not imply, of course, that it must always lead to an actual fall in prices. It may merely slow down an existing inflationary process for some time, without actually reversing it.

It is also of crucial importance that the currency be devalued to its equilibrium level; the removal of import restrictions is otherwise impossible, or feasible only to a limited extent. In the most recent French devaluation, for example, the new rate clearly did not attain the equilibrium level, as indicated by the discount of almost 10 per cent on the new exchange rate at which francs continued to be traded in the free currency markets of the world.

As in the case of tariffs, our contention is further strengthened if currency overvaluation together with the accompanying trade restrictions is not confined to one single country only. Multilateral imposition of controls, the equivalent of retaliatory tariffs, is bound to reduce real income all around.

III

A few comments on the differences between various restrictive policies are in order. If the government uses quotas, and import licenses are auctioned off, the premiums paid by importers are the exact equivalent of the import duty that would have achieved the same degree of restriction. If import licenses are allocated without, or against an only insignificant fee, and if no price controls exist, the receivers in effect reap windfall profits of the same magnitude. It would be hasty, however, to conclude that such an arrangement must be an importer's paradise. He earns the higher unit profits on a reduced volume of trade, and his income may well fall below the level it would attain under unrestricted trade.

If price controls are in effect, it may appear at first sight that our conclusions have to be substantially modified, if not reversed.

Price controls may concern imported commodities only; such arrangements have occasionally existed. But since the national product is, in plain Hollywoodese, a many-splendored thing, the excess purchasing power will merely spill over into other markets. The price rise in the other sectors can be confidently expected to neutralize the artificial stability of import prices.

Where rationing and price controls have been all-embracing, on the other hand, a black market has hardly ever been absent. And an economist computing a meaningful price index, however indignant a moralist he may be, has to take black market prices and volumes traded into consideration. The outcome may be hardly any different from our general conclusion.

The Utopian society in which over-all price controls and rationing are indeed fully effective, forces a ceiling on its consumption level at the same time that it succeeds in keeping the price level down. Under such conditions, the price level may indeed rise when price as well as exchange controls are removed after depreciation. But if the move towards freer trade raises the potential welfare level, i.e., when it produces a more abundant supply of commodities, it is difficult to see why a fetish should be made out of a lower price level at a forcibly reduced standard of living.

Our general conclusions will be reinforced by the absence of perfectly competitive conditions in the real world. Existing monopolies can only be weakened by the rise in the volume of foreign trade that the adjustment of exchange rates to their equilibrium levels makes possible. The closer approach to the competitive ideal will be one further factor contributing towards a fall in the domestic

price level. This factor will be all the more important the smaller a country is.

It is, of course, not permissible to compare the present pattern of overvalued exchange rates and partial financing of trade deficits by unilateral government grants with a situation in which all exchange rates are at the free competitive level, but all flows of government grants have ceased. Two situations can be objectively compared only when all capital transfers⁷ remain at the same level. This has been implicitly assumed so far. The Machiavellian view may be entirely justified that foreign aid might no longer be forthcoming as soon as the absence of visible balance-of-payments disequilibria brings about a revision of official government policies under the pressure of public opinion. But this is a question outside the realm of economic analysis.

However, the increase in allocative efficiency may well be significant enough to dwarf the reduction in the commodity intake which the removal of a deficit in its balance of payments superficially appears to force upon a country. It is entirely possible, therefore, that a depreciation may *both* wipe out the deficit (and thus eliminate the need for assistance from abroad or the drain on the country's gold and exchange reserves) *and* cause a deflationary impact on prices even in the absence of deflationary monetary and fiscal policies. The Austrian depreciation of 1953 may serve as an example that such an outcome is by no means wishful thinking.

Any residual disbelief in this proposition will be dispelled if the reader bears in mind again that a deficit may be eliminated not only by an increase in exports or a fall in imports, but also by an increase of *both* exports and imports, provided only that the value of the former rises by more than that of the latter. This is the most likely outcome whenever restrictions on trade are in effect before the depreciation. It is, furthermore, not only the aggregate value, but also the *composition* of the import bill that matters.

Our analysis is by no means intended to suggest, on the other hand, that a cancellation of foreign aid programs after an adjustment of exchange rates would be desirable. They should be regarded as a natural extension to the international sphere of basic principles of public finance long accepted for domestic use in all Western countries.⁸ Given the willingness of richer countries to provide such aid, all

7. See note 7, p. 275 for the definition of capital transfers used in the present paper.

8. This is Gunnar Myrdal's main thesis in his recent book, *An International Economy, Problems and Prospects* (New York: Harper, 1956).

evidence appears to support the thesis that not only the paying, but also the receiving countries would derive substantial benefits from a return to exchange rates that are compatible with a freer flow of international trade.

EGON SOHMEN.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

THE BUSINESS CYCLE IN THE POST-WAR WORLD: A REVIEW*

By JAMES TOBIN

The International Economic Association assembled in Oxford in September 1952 a group of distinguished students of economic fluctuations from all over the world. The papers of the conference were of two types: highly informative analytical narratives of post-war economic developments in eight countries and contributions to the theory of short-run economic fluctuations. The country papers make up Part I of the volume, the theoretical papers Part II; these are followed by A. D. Knox's very readable summary of the give-and-take in discussion among the authors and other conference participants. D. H. Robertson begins the volume with an Introductory Note, which is such a penetrating and balanced review of the book, written in the always incomparable Robertsonian style, that any subsequent reviewer is bound to feel redundant and heavy-handed. (Sir Dennis has also, by the way, contributed a poem, "The Non-Econometrician's Lament," which by itself is worth the price of the book.)

The ill-chosen title of the conference sent the participants off on a search for "the business cycle" in economic developments since 1945. Wherever the hunters looked — United States, United Kingdom, Sweden, Denmark, France, Germany, Italy, or Japan — their quarry was somewhere else. A tacit premise of this international quest seemed to be that The Trade Cycle did exist before the war. If by "The" some uniqueness in the cause and mechanism of economic fluctuation is implied and if "Cycle" means any regularity in the duration and amplitude of business ups and downs, the premise may be doubted. But even if it is granted, it is hardly surprising that the cycle should prove elusive in the immediate postwar years, dominated as they were by the effects of the war and of its differential impacts on national economies. No doubt "the cycle" has lost most of its dictionary meaning for many economists, who use the term simply as a synonym for all those economic fluctuations that cannot be traced either to the annual rotation of the earth

* Proceedings of a Conference held by the International Economic Association. Edited by Erik Lundberg, assisted by A. D. Knox (London: Macmillan, 1955).

around the sun or to gradual decade-to-decade changes in population and other basic resources. Even on this dilute interpretation, it was scarcely to be expected that fluctuations in 1945-52 would belong to the same family as those before either world war. For one thing, virtually all the movement was upwards; the studious attention lavished on the short-lived shallow dip in 1948-49 is mainly indicative of its scarcity value.

Nevertheless, postwar economic experience, as reviewed at the conference, does contain important general lessons for the theory of business fluctuations, lessons that have significance beyond the special circumstances of those years. Underlying all national economic developments immediately following the war were the adjustments required to restore some kind of normal balance between the two main vehicles for private holding of wealth, liquid claims against governments and financial institutions, on the one hand, and physical capital, either consumers' or producers' goods, on the other. The legacy of the war and of the preceding depression was, almost everywhere, a shortage of stocks of goods relative to the population of consumers and workers, accompanied in many countries by a backlog of technological innovations yet to be applied in civilian production. The marginal productivity of producers' capital, and the marginal utility of the services of durable consumers' goods were high. At the same time, the financing of the war had left large accumulations of low-yielding liquid claims in the hands of savers. The lowness of the nominal yield on these claims was compounded, in many countries, by widespread anticipation of decline in their purchasing power.

The immense disproportions between liquid claims and physical goods in the "portfolios" of owners of wealth, from the most humble peasant to the largest corporation, set the stage for the dramatic struggles of wealth owners to increase their holdings of physical capital and to reduce their relative holdings of monetary and quasi-monetary assets. These struggles set in motion a number of adjustment mechanisms: (a) Over time, the production of new durable goods adds to their supply and — unless human population and capital-using technological advances accumulate even faster — reduces their yield. (b) The nominal yields on liquid claims rise, increasing their attractiveness as stores of wealth. During the period surveyed, however, significant increases in interest rates were avoided by deliberate policy in the United States, United Kingdom, and Sweden. (c) Inflation lowers the real value of liquid claims, diminishing the total value of a given nominal supply relative to the stock

of capital. If half-hearted and transient gestures at price and wage controls are ignored, this classic mechanism was permitted to operate in all the economies studied except the United Kingdom, Sweden, and Denmark. (d) The supply of liquid claims is drastically reduced by partial confiscation of private holdings. A radical "currency reform" was possible only in occupied Germany and Japan. (e) Government budget surpluses diminish the net supply of liquid claims in the hands of the public. In occupied Japan, this policy was imposed on the government by the omnipotent Joseph Dodge. (f) Balance-of-payments deficits, due to the appetite for imported goods, reduce the supply of liquid claims, both international reserves and the corresponding domestic monetary assets.

As William Fellner notes in his illuminating and judicious report on the United States, there are a number of alternative measures of the relative liquidity of an economy. All of these would show excessive liquidity at the end of the war, and a reduction in liquidity thereafter. Consequently the theoretical question of the appropriate measure can be dodged for the period under study, and Fellner dodges it. To illustrate the course of relative liquidity in the United States, Fellner uses a common variant of the Cambridge K , the ratio of the money supply (currency plus demand and time deposits in commercial banks less U. S. government holdings) to private gross product (Gross National Product less U. S. government purchases of goods and services.)

A similar emphasis on liquidity, similarly measured, is offered by Sir Dennis Robertson's observation, "It is surprising for how many countries, . . . the tale of the post-war expansion can be told in terms of a prolonged and ultimately fairly successful struggle on the part of K to return to something like a pre-war level."¹

In less dramatic times, alternative measures of relative liquidity may not always move in the same degree or even in the same direction. If correct lessons for more normal times are to be drawn regarding the importance of relative liquidity, the theoretical issue of the relation of liquidity to economic activity must be squarely faced. This review is not the place for the development of such a theory, but several related points can be made:

(1) The conventional definition of the money supply is not an adequate concept of the supply of liquid claims. There is a whole spectrum of assets that are more or less close substitutes for com-

1. A recent and detailed interpretation of postwar European monetary experience in these terms is given by Robert Triffin, *Europe and the Money Muddle* (New Haven, Yale University Press, 1957).

mercial bank deposits and currency as vehicles for holding wealth. Government bonds and savings accounts are conspicuous examples. Excess liquidity, as a factor in the demand for goods, is not to be measured merely by the supply of deposits and currency. The whole internal government debt, whether monetized or not, must be considered. In addition, both banks and other financial institutions whose obligations are liquid claims closely substitutable for money add to the relative liquidity of the economy when they lend to individuals or firms to enable the borrowers to increase their holdings of houses, inventories, or other forms of capital.

(2) The supplies of liquid claims of various kinds must be compared not only with some measure of national income but with the size of the capital stock. A normal value of Marshallian K is not as strategic as a normal balance between the two main forms in which the community can hold wealth, liquid claims against the government and equity in real capital.

(3) Monetization of government debt has much less economic significance than monetization of capital. When banks buy government bonds and create deposits, they merely change in a slight degree the form of the liquid claims held by the public. When banks and other financial institutions need credit in order to lend to entrepreneurs who invest in inventories, the community's demand for capital is increased and at the same time the appetite of more cautious members of the community for liquid assets is satisfied by the corresponding deposits. Since 1952 in the United States there has been a further decline in the Marshallian K ; the stock of money, conventionally measured, has remained constant and national product has grown. But it would be a mistake to regard the activities of the banking system and of other financial institutions during this period as neutral or deflationary. Lending to business firms and home-owners expanded at the expense of lending to the government. Monetization of capital increased, while some of the national debt was "demonetized" and had to be placed directly rather than indirectly with corporations, institutions, and individuals desiring to hold liquid claims.

Relative liquidity considerations play a strategic role in most of the country narratives, but the papers are all more or less handicapped by the lack of appropriate analytical tools and concepts. In this respect the book is another reminder of the customary lag of economic analysis behind events. The tools built for the thirties are as inadequate for the forties and fifties as the tools of the teens and twenties were for the Great Depression. One way to charac-

terize the difference between the eras before and after World War II is to say that interdependences between stocks and flows, which during the depression were almost wholly dominated by relations between flows and flows, came to be of central importance after the war. Many economists have appreciated the nature of the change but, lacking more modern tools, have had to fall back on the oversimple concepts of pre-Keynesian monetary theory. Examples in the present book include not only Fellner and Robertson but also Preiser and Krelle, who present an econometric model² of economic developments in West Germany. In their model total expenditure is obtained by multiplying the supply of money by an income-velocity which is assumed to be influenced only by seasonal factors, price expectations, and accidental shocks.

Other analysts have attempted to interpret postwar developments in terms of the calculus of interrelations among flows, exemplified by "inflationary gap" analysis. The inadequacies of these interpretations are most apparent in the accounts of economic events and government policy in the United Kingdom, Sweden, and Denmark. The authors — E. A. G. Robinson, Erik Lundberg, and Jorgen J. Pedersen respectively — were uneasy over the analyses of demand for and supply of National Product on which economic planning and policy in their countries were based. But their own interpretations use the same framework and suffer likewise from the incidental position assigned in the analysis to stocks — as opposed to flows — of liquid claims and capital goods. In the case of Japan, the "Dodge-line" policies of the government cannot be charged with neglecting the monetary determinants of demand, but Shigeto Tsuru's narrative is handicapped by an excessive preoccupation with the flows of national income components. For example, Tsuru emphasizes the disinflationary consequences of budget surpluses on the contemporaneous income flows but neglects the lasting disinflationary effects of the reductions in supplies of liquid claims accomplished by these surpluses.

The story of France is told by André Piatier in terms that emphasize quite clearly the importance of the initial imbalance

2. The parameters of this model were not estimated by techniques that permit any tests based on criteria of goodness-of-fit. They were rather estimated by a trial-and-error process designed to obtain for the fourteen quarterly periods of observation residuals that corresponded in sign and order of magnitude with a priori judgments about the nature of external disturbances. Since the exact size of these external influences can also be regarded as parameters to be estimated, it is clear that the authors had very few, if any, degrees of freedom on which to base confidence in their model.

between supplies of liquid claims and the stock of goods and the mechanisms by which this imbalance was corrected. The Italian economy's course during 1945-52 is described in interesting fashion by Pasquale Sarraceno. Italy is an exception to the inflationary full employment conditions that characterized the other countries. Here the years after 1947 exhibited a surprising combination of events: structural unemployment, an extremely high marginal productivity of capital, an increasing supply of money, and price stability. No adequate explanation of Italian experience emerges from Sarraceno's account; a comparative analysis of France and Italy would be highly instructive.

If the country papers of Part I show that the analytical tools of economic theory lag behind economic events, the theoretical contributions of Part II exhibit the lag even more clearly. Liquidity factors, which occupied the center of the stage in the postwar dramas of most national economies, are relegated to minor roles in the models of the theorists.

A happy exception is provided by the stimulating paper of J. J. Polak, which attempts to support the thesis that there is a cycle inherent in the mechanisms of countries' responses to changes in external demands and supplies. Polak argues that there is more to international propagation of cycles than simple duplication, with magnification, of the cyclical movements of major countries, such as postwar United States. Very commendably, Polak attempts to subject this thesis to statistical tests, but the results can be no more than suggestive because too few years and countries are available for testing. But whether or not Polak has detected an international response-mechanism cycle, the striking thing about his paper is that he has two explicit models, one for the prewar mechanism, the other for the postwar mechanism. The difference between the models is, on an international scale where nations and governments are the units of behavior, essentially the same difference that distinguishes prewar and postwar internal economic environments. His prewar model runs in terms of relations among flows: for each country, exports determine income by a multiplier process and income determines imports. The model is appropriate to a world in which a nation's deficit on international accounts is not limited by its supply of liquid international reserves. In Polak's postwar model, the primary determinant of imports is not the country's income but its reserve position, the cumulative result of its balance-of-payments deficits or surpluses.

Nicholas Kaldor discusses the international propagation of busi-

ness fluctuations largely in terms of a more elaborate version of Polak's prewar model. His purpose is more normative than analytical. He is concerned with measures by which countries may protect themselves from fluctuations emanating from other economies. For this purpose he is led to distinguish between countries (i.e., the United States) with ample reserves and those with limited reserves. If the United States will not, by its own good behavior, stabilize its supply of dollars to the rest of the world, Kaldor must recommend that countries short of reserves behave in the manner assumed by Polak's postwar model. Since, according to Polak, that response mechanism also produces a cycle, it is not clear that Kaldor's prescription really accomplishes his objective.

Richard Goodwin presents an elegant and ingenious model of the relationship between cycles and economic growth. Goodwin's models always command admiration for the clarity and simplicity with which they capture the significant implications of fundamental features of the capitalist process, in particular the role of innovations and the dependence of investment on the existing capital stock. Goodwin's models are stripped, but of nonessentials rather than essentials. His work draws inspiration jointly from Schumpeter and Keynes, whose visions he endeavors to formalize and to blend. His "flexible accelerator" is a substantial improvement over the mechanical capital-output ratio that characterizes the popular models of Harrod and Domar, and it is no credit to the profession that Goodwin's work has received little attention in comparison with the overworked rigid accelerator. In the present volume his previous model of the flexible accelerator is combined with a ratchet-effect secular consumption function and a constant flow of potential innovations.

Goodwin's ingenuity makes it possible to go a long way in describing the twin processes of growth and cycles in terms of aggregate capital requirements alone. But not far enough. Prices, profit rates, interest rates, liquidity supplies, and bank credit are conspicuously absent from Goodwin's theory, as they are from simpler versions of the dynamics of capital accumulation. The absence of monetary and price phenomena is especially conspicuous in a volume where these variables play such strategic roles in all the country narratives. An accelerator model, whether flexible or rigid, cannot explain inflation and deflation, nor can it allow the lengths of booms and depressions to depend on fiscal and monetary events and policies. Goodwin's model illuminates half of an important truth: whether the economy is booming or depressed depends on whether capital is short or redundant. It misses the equally important other half of this truth: the adequacy or redundancy of the stock of capital at a

given income level is not a knife-edge question, but depends on the relative supplies and yields of capital and alternative means of holding wealth.

R. A. Gordon's informed and imaginative discussion of investment opportunities in the United States suffers from the same neglect of competing vehicles for holding wealth. For Gordon, as for Goodwin, the stock of investment opportunities can be gauged wholly from technological considerations determining the adequacy of the size and composition of the existing capital stock to the size and composition of national product. Gordon's emphasis on the composition, as well as the size, of national capital and product is highly fruitful. Aggregation conceals the fact that capital is accumulated in specific forms; railroad track is useless in the aircraft industry. Gordon's projections have stood very well the test of time; the recession of 1956-58 confirms his 1952 estimate that (assuming the military program tapers off) it would not be easy for the economy to generate a full-employment level of investment in the late 1950's.

The remaining theoretical paper, by Lawrence R. Klein, is principally an exposition of the methodology of formulating and statistically estimating systems of equations representing the short-run determinants of over-all economic activity. Much of the exposition is familiar ground to anyone who has followed Klein's pioneering work in the past. His present paper shows a healthy concern for augmenting the too limited evidence of aggregate time series with cross-section statistics, and contains useful suggestions for integrating the results of surveys of investment and consumption intentions into econometric forecasting models. The model Klein presents as illustrative of a model to be statistically estimated is a somewhat disaggregated neo-Keynesian system of nineteen equations, in which liquidity supplies are rightly permitted to play a significant role. In comparison with earlier econometric cycle models proposed by Klein, this one has a much greater degree of year-to-year persistence built into it. The values of variables in the current year are made to depend on, among other things, their previous year values. Built-in persistence has the advantage of avoiding the gross errors that can result from attempting, so to speak, to reconstruct the economy from scratch every year. But it makes the detection of turning-points more difficult; in the extreme case, the naive prediction that tomorrow will be like today would generally be correct but would be a useless forecast. Whether Klein has now struck the right compromise remains to be seen.

JAMES TOBIN.

INVESTMENTS IN UNITED STATES GOVERNMENT SECURITIES BY NONFINANCIAL CORPORATIONS, 1952-56: COMMENT

By COLIN D. CAMPBELL

In "Central Banking and Money Market Changes,"¹ Hyman Minsky finds that financial institutions and money market usages change so as to counteract a tight money policy. As one example, he states that high interest rates have influenced nonfinancial corporations to hold more and more of their cash in short-term U. S. government securities and have thus freed bank resources to finance other activities. Although a shift in the ownership of short-term federal securities from commercial banks to nonfinancial corporations could counteract a tight money policy, such a shift has probably not occurred. Minsky's interpretation of the statistical data on the ownership of federal debt from December 1952 to June 1956 is very questionable.

To show that there has been a shift of short-term federal debt from commercial banks to nonfinancial corporations, Minsky presents data from the Treasury Survey of Ownership showing (1) that commercial bank holdings of marketable issues maturing within one year declined sharply from December 1952 to June 1956 and (2) that during this period such investments by "other investors" including nonfinancial corporations expanded \$5.7 billion. The series on "other investors" is an inappropriate measure of the changes in the holdings of federal securities by nonfinancial corporations because there are other important groups in this category — state and local governments, individuals (including partnerships and personal trust accounts), "miscellaneous investors" (including savings and loan associations, dealers and brokers, foreign accounts, corporate pension trust funds, and nonprofit institutions), and those banks and insurance companies not reporting in the Treasury survey. In addition, Minsky's estimate of changes in the amount of short-term federal securities owned by nonfinancial corporations omits Treasury savings notes, a nonmarketable issue.² From 1941 to April 1956 these notes were widely used by corporations as liquid assets, but in 1953 the

1. This *Journal*, LXXI (May 1957).

2. Treasury savings notes could be used without notice for payment of taxes two months after issue date at purchase price plus accrued interest, and they were redeemable for cash four months after issue.

Treasury suspended sales of Treasury savings notes, and none was outstanding in June 1956. The Treasury survey shows that in December 1952 "other investors" held \$5.7 billion of these notes.³ If Treasury savings notes and marketable U. S. government securities maturing within one year are added together, the total amount of these securities owned by "other investors" was approximately the same in June 1956 as in December 1952. In spite of this, during this period the amount of these types of securities owned by commercial banks declined \$9.6 billion. This was primarily the result of a decrease of \$4.1 billion in the amount of these securities issued by the Treasury and the addition of \$5.5 billion of such securities to the portfolios of the Federal Reserve Banks. Although point-of-time comparisons are tricky and shifts of only a few months give different results, the above analysis shows some of the factors that are involved.

Minsky could have obtained a better measure of changes in the amount of short-term federal securities owned by nonfinancial corporations from Treasury estimates of their total holdings.⁴ This is because the bulk of their holdings consists of short-term issues. Persons closely associated with the government securities market generally believe that nonfinancial corporations invest primarily in short-term securities.⁵ Also, a study of the composition of the investments of the 100 largest nonfinancial corporations in 1951 shows that the bulk of their holdings consisted of short-term issues.⁶ Table I shows that the total amount of federal securities owned by nonfinancial corporations declined \$2.5 billion from December 1952 to June 1956.

Contrary to what Minsky says, the amount of U. S. government securities owned by nonfinancial corporations does not appear to be closely related to changes in interest rates. The following charts show that from 1951 to 1956 there appears to have been a slightly downward long-run trend in holdings of federal securities by nonfinancial corporations and an upward long-run trend in short-term

3. *Treasury Bulletin*, Mar. 1953, p. 38.

4. The Treasury publishes monthly estimates of the total amount of the national debt owned by nonfinancial corporations, state and local governments, individuals, and "miscellaneous investors." No regular data on the kinds of federal securities owned by these groups are available. The estimates of the total amount of federal securities owned by nonfinancial corporations are based on quarterly data compiled by the Securities and Exchange Commission and published in its report on Current Assets and Liabilities of Corporations. See *Federal Reserve Bulletin*, XLIII (May 1957), 558.

5. See M. Nadler, S. Heller, and S. S. Shipman, *The Money Market and Its Institutions* (New York, 1955), pp. 268-69.

6. J. S. Sprowls, "Short-Term Investment Practices of Large Non-Financial Corporations," Table 2, MBA Thesis, University of Pittsburgh, 1953.

TABLE I
OWNERSHIP OF UNITED STATES GOVERNMENT SECURITIES, 1952-1956
(In billions of dollars)

End of Month	Total Public Holdings ¹	Commercial Banks	Other Financial Institutions ²	Nonfinancial Corporations	State and Local Governments	Savings Bonds	Individuals Other Securities	Miscellaneous Investors ³
Dec. 1952	196.9	63.4	25.6	19.9	11.1	49.2	16.0	11.7
Dec. 1953	201.0	63.7	25.0	21.6	12.7	49.4	15.4	13.2
Dec. 1954	204.3	69.2	23.8	19.2	14.4	50.0	13.7	13.9
Dec. 1955	204.3	62.0	22.8	23.3	15.1	50.2	15.4	15.6
June 1956	195.5	57.1	21.7	17.4	15.7	50.3	17.2	16.2

Source: *Treasury Bulletin*, August 1957, p. 40.

1. U. S. government securities held outside Federal Reserve Banks and U. S. investment accounts.

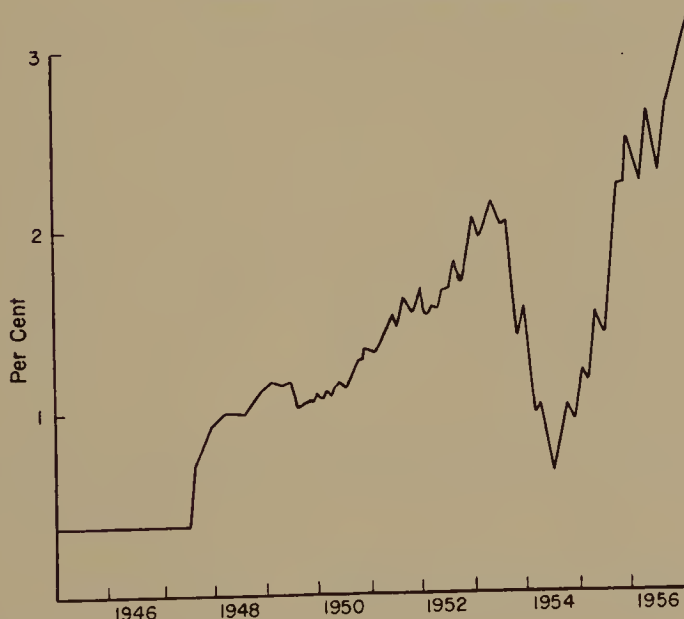
2. Includes mutual savings, banks and insurance companies.

3. Includes savings and loan associations, dealers and brokers, foreign accounts, corporate pension trust funds, and nonprofit institutions.

U. S. GOVERNMENT SECURITIES OWNED BY NONFINANCIAL CORPORATIONS,
JUNE AND DECEMBER, 1945-1956

Source: *Treasury Bulletin*

YIELDS ON U. S. TREASURY BILLS



Source: Federal Reserve Charts.

interest rates. Also, in several years changes in such holdings were in the opposite direction from changes in short-term interest rates. From December 1946 to December 1948 the amount of federal securities owned by nonfinancial corporations fell even though the bill rate rose sharply. In 1949 their holdings increased substantially even though the bill rate fell slightly. In 1952 their holdings dropped even

though the bill rate rose. During some years regular fluctuations have resulted from the accumulation of federal securities prior to corporation income tax payment dates. In these years the amount of these investments fell during the first half of the year and rose in the second half of the year. Investments by nonfinancial corporations in federal securities are not sensitive to changes in interest rates, probably because the principal purpose of such investments is to provide liquidity.

From December 1952 to June 1956 commercial banks, mutual savings banks, and insurance companies obtained over \$10 billion in loanable funds through the decline in their holdings of U. S. government securities, even though commercial bank holdings rose sharply in 1954. These securities were absorbed by state and local governments, individuals, and "miscellaneous investors" rather than by nonfinancial corporations. Table I shows that investments in U. S. government securities by state and local governments expanded \$4.6 billion during this period. Although the principal reason for this expansion is the growth in their retirement systems, which invest primarily in long-term issues, state highway and construction funds have purchased large amounts of federal securities in recent years. Since these funds are usually the proceeds of state bonds sold for construction purposes, they can be invested only in short-term issues.⁷

Table I also shows that the amount of federal securities owned by "miscellaneous investors" expanded \$4.5 billion. Foreign accounts probably acquired more federal securities during this period than any of the other groups in this category. From December 1952 to June 1956 investments of foreign banks and official institutions in bills and certificates increased \$2.4 billion.⁸ During these years foreign countries and international institutions have held approximately half of their dollar balances in U. S. government securities, and favorable balances of trade have caused their total dollar balances to increase more than one-third.⁹ Also, during this period the amount of U. S. government securities owned by savings and loan associations increased \$.9 billion, and those owned by corporate pension trust funds by \$.5 billion.¹ Both of these groups invest primarily in intermediate and long-term issues.

7. C. D. Campbell, "Investments in United States Government Securities by State and Local Governments," *National Tax Journal*, X (Mar. 1957), 85.

8. *Treasury Bulletin*, Mar. 1953, p. 69; *ibid.*, Aug. 1956, p. 67.

9. "International Gold and Dollar Flows," *Federal Reserve Bulletin*, XLIII (Mar. 1957), 249.

1. For data on the assets of savings and loan associations, see *Federal*

The amount of U. S. government securities other than savings bonds owned by individuals rose \$1.2 billion during this period. This estimate is a residual figure after estimates have been made for other ownership groups, and little is known about the composition of these investments.

Recent changes in debt ownership are not as alarming as Minsky expected primarily because a large part of the funds used for expanding the loans of banks and other financial institutions has been obtained through a shift of *intermediate* and *long-term* issues — either through direct sale or Treasury refinancing operations — to state pension funds, corporate pension funds, and savings and loan associations. As Minsky suggests, if nonfinancial corporations substituted *short-term* securities for bank deposits, and financial institutions selling short-term securities expanded their loans, the effect on the supply of loanable funds or on total spending would be expansionary. This is because the change in the composition of the current assets of nonfinancial corporations would probably not affect their spending. To the extent that financial institutions have shifted short-term federal securities to foreign banks or to state construction funds, the effect on the supply of loanable funds could be just as expansionary as shifting such securities to nonfinancial corporations. A crucial factor in this expansionary process is usually the willingness of banks to hold smaller amounts of secondary reserve. However, because bank portfolios of U. S. government securities become more liquid as outstanding issues approach maturity, banks are frequently in a position to dispose of some of their short-term issues without reducing secondary reserve.

COLIN D. CAMPBELL.

DARTMOUTH COLLEGE

Reserve Bulletin, XLIII (May 1957), 549, and for estimates of investments in federal securities by corporate pension trust funds, see *Treasury Bulletin*, Mar. 1954, p. 30; *ibid.*, Sept. 1956, p. 35.

REPLY

By HYMAN P. MINSKY

In my article "Central Banking and Money Market Changes," I discussed two recent institutional changes: the growth of the federal funds market and the development of a new financing technique for government bond houses, the sales and repurchase agreements between the bond houses and nonfinancial corporations. In com-

menting upon the second change, I drew attention to a number of alternative forms by means of which nonfinancial corporations can achieve liquidity: demand deposits, short term Treasury debt, sales and repurchase agreements with bond houses and loans to other financial intermediaries such as consumer credit houses. Observing that short term debt holdings by commercial banks markedly decreased while the holdings of other investors (which include nonfinancial corporations) increased in the period December 31, 1952 to June 30, 1956, I imputed these observed changes to the rise in interest rates over this period.

In his note Professor Campbell correctly points out that the class "other investors" includes important groups, such as state and local governments, individuals, etc., in addition to nonfinancial corporations. He argues that it would have been better to use the available data on the "Ownership of United States Government Securities, 1952-1956" (his Table I), even though "no regular data on the kinds of federal securities owned are available." Using these data, which show that holdings by nonfinancial corporations of U. S. government securities have gone down slightly in this period, he reaches two conclusions of interest:

(1) "Investments by nonfinancial corporations in federal securities are not sensitive to changes in interest rates. . . ."

(2) "Recent changes in debt ownership are not as alarming as Minsky expected primarily because a large part of the funds used for expanding the loans of banks and other financial institutions has been obtained through a shift of *intermediate* and *long term* issues . . . to state pension funds, corporate pension funds, and savings and loan associations."

Campbell errs in interpreting my argument with regard to the holdings of short term government debt by the other category as implying that the amount of U. S. government securities owned by nonfinancial corporations is closely related to the interest rate. What I argued was that the form in which *liquidity* is held is sensitive to the quality of the available assets and the relative interest rates. As long as the interest rate on demand deposits is zero, any increase in interest rates increases the attractiveness of alternative acceptable assets, and hence a substitution of such assets for demand deposits will take place.¹ However, nonfinancial corporations also hold government securities as investments (that is, these assets are superfluous to the current operations of the firm), and a tightening money market during a boom would make firms with favorable opportunities draw upon their superfluous assets to finance expansion. The decrease in

the holdings of government securities by nonfinancial corporations (see Campbell's chart) immediately after World War II can be explained in this manner. The observed changes in the total holdings of government debt by nonfinancial corporations, to which Campbell refers, is consistent with a running down of investment holdings of longer term debt at the same time as short-term government debt is substituted for demand deposits.

Campbell is correct in pointing out that I neglected Treasury savings notes. However, he does not mention that since 1950 the acceleration of corporate income tax payments has been operating to decrease the need of corporations for liquidity. Whereas in 1950 corporate income tax payments lagged by one year behind the earning of income, the acceleration of payments will reduce this lag to six months by 1960. On June 30, 1952 corporations which kept their books on a calendar year basis still owed 30 per cent of the tax on their 1951 income, and on December 31, 1952 none of the tax on 1952 income had been paid. On June 30, 1956 such corporations had paid all of the tax on 1955 incomes, and by December 31, 1956 they had paid 20 per cent of the tax on their estimated 1956 income.² My estimate is that between 1952 and 1956 the need of corporations for liquidity on account of their income tax liability was reduced by 20 to 25 per cent; that is, the income tax liability of corporations on December 31, 1956, which was \$16.6 billions,³ was \$4 to \$5 billions less than it would have been if the payment schedule had not been changed since 1952.

Treasury savings notes were a liquid asset particularly suited for corporations with an income tax liability. Even though in the period 1952-56 approximately two-thirds of the redemptions were for cash (in sharp contrast with earlier experience), the elimination of the Treasury savings notes and the reduction in corporation liabilities on account of the acceleration of tax payments roughly offset one another. Hence I would argue that the existence of nonmarketable Treasury savings notes in 1952 can be ignored in considering the effects of interest rate changes upon the form in which nonfinancial corporations achieve liquidity.

1. Incidentally Campbell is aware of this in so far as state and local governments are concerned: see C. D. Campbell, "Investments in United States Government Securities by State and Local Governments," *National Tax Journal*, X (Mar. 1957), 86.

2. U. S. Treasury Department: *Annual Report for the Fiscal Year Ending June 30, 1954*, pp. 285-86.

3. U. S. Department of Commerce: *Statistical Abstract of the United States, 1957*, No. 592, "Current Assets and Liabilities of United States Corporations, 1945-1956," p. 486.

Over this period the government debt holdings of commercial banks decreased by 6.3 billions, with an essentially unchanged volume of government debt outside government trust funds and Federal Reserve Banks. As a result, the lending ability of commercial banks was increased by operations which are equivalent to the banks selling government debt to other holders. Contrary to what Campbell concludes in his final paragraph, the monetary effect of such transactions is independent both of the particular category in the other investors group that acquired the government debt, and of the dating of the debt.

If current savings were used to purchase the government debt which was shifted from the portfolios of commercial banks, then the current savings would appear as a supply of funds in the form of an increase in the lending ability of commercial banks. On a net basis the transactions would not be inflationary. On the other hand, if such sales were effected by using previously accumulated demand deposits, the resulting increase in the lending ability of commercial banks does not offset some current savings. Such a development is inflationary. The effects are independent of the dating of the government debt shifted, and the tendency to identify the monetary effect of a shift in government debt with the dating of the securities is erroneous. For example, Campbell notes a net purchase of intermediate and long term issues by savings and loan associations. If the increase in savings and loan deposits is a result of a shift from demand deposits to interest-earning time deposits, then the net acquisition of government debt by savings and loan associations at the same time as commercial bank deposits were decreasing is inflationary.⁴ On the other hand, the shift of government debt to the various pension funds, which can be considered as savings intermediaries, is not inflationary.

HYMAN P. MINSKY.

BROWN UNIVERSITY

4. In the United States, commercial banks have both demand and time (savings) deposits. An increase in commercial bank time deposits can also be the result of savings (in which case it is actually deflationary due to the member bank reserve requirements against savings deposits) or of a shift in liquidity (in which case it is inflationary). Conceptually commercial banks can be departmentalized, and an increase in their holdings of long and intermediate term government debt in response to an increase in savings deposits could occur at the same time as their holdings of short term government debt decreased in response to their "unsatisfactory" returns.

TARIFFS AND THE BALANCE OF PAYMENTS : COMMENT

By HARRY G. JOHNSON

In his article on "Tariffs and the Balance of Payments,"¹ Dr. Ozga shows that the imposition of a tariff may lead to a deficit in the balance of payments, instead of a surplus. While Ozga's analysis is correct, his statement of his result may be somewhat misleading, since it tends to imply that the result is due to the imposition of the tariff rather than to the expenditure policies assumed to accompany the tariff; and his explanation of the reasons for the result is not entirely clear or satisfactory. The following alternative treatment of the problem may therefore be of some assistance in understanding his conclusions.

Suppose trade between *A* and *B* is in equilibrium with a deficit for *A*. The deficit implies an excess of expenditure over output in *A*, and vice versa in *B*; as a first step towards correcting it, *A* must deflate expenditure and *B* must inflate expenditure by the amount of the deficit. But the combined deflation-inflation is likely to alter the distribution of demand between the two countries' goods, so that some further measure will be necessary to readjust demand to supply for the separate goods. If the sum of the marginal propensities to spend on imports is less than unity, the combined deflation-inflation would leave a deficit for *A*, an excess supply of *A*'s goods, and an excess demand for *B*'s goods, at the initial prices; to restore equilibrium, demand must be diverted from *B*'s goods to *A*'s goods, which requires a fall in the relative price of *A*'s goods, a depreciation of *A*'s currency (assuming elasticities are favorable) or (as in Ozga's model) an increase in *A*'s tariff. Conversely, if the sum of the marginal propensities to spend on imports is greater than unity, the combined deflation-inflation would leave a surplus for *A*, an excess demand for *A*'s goods, and an excess supply of *B*'s goods, at the initial prices; to restore equilibrium, demand must be diverted from *A*'s goods to *B*'s goods, which requires a rise in the relative price of *A*'s goods, an appreciation of *A*'s currency, or a reduction in *A*'s tariff. It follows that in this case a higher tariff goes with a larger deficit, as Ozga has found. But it is misleading to imply that it is the higher tariff which causes the deficit: the higher tariff merely

1. S. A. Ozga, "Tariffs and the Balance of Payments," this *Journal*, LXXI (Nov., 1957), 630-38.

happens to be the condition of equilibrium between demand and supply for the individual countries' outputs, when national expenditures are so altered that the deficit is larger, and it remains true that the tariff by itself tends to improve *A*'s trade balance.

The explanation of the point can also be put another way, paralleling the argument of Ozga's Section IV. The imposition of an (expenditure-compensated) tariff in *A* diverts *A*'s demand from its imports towards its exports — the substitution effect — thus creating an excess demand for *A*'s goods, an excess supply of *B*'s goods, and a balance-of-payments surplus for *A*. To restore equilibrium without changing prices in *B*, the excess demand and supply must be eliminated by deflation of expenditure in one country and inflation of expenditure in the other. If the sum of the marginal propensities to spend on imports is less than unity, deflation in *A* and inflation in *B* will do the trick; in so doing it will increase *A*'s surplus still more. But if the sum of the marginal propensities to import exceeds unity, inflation in *A* and deflation in *B* is required; this implies a worsening of *A*'s balance of payments to set against the improvement due to the tariff. Further, since the effect of a unit inflation in *A* and deflation in *B* is to worsen *A*'s balance by the sum of the marginal propensities to spend on imports, while it reduces excess demand for *A*'s goods only by this sum minus one,² the removal of the excess demand due to the tariff by a combined inflation-deflation will worsen *A*'s balance more than the tariff initially improved it, so that the net result will be a deterioration in *A*'s balance of payments.³ Again, the deterioration should be attributed to the expenditure changes accompanying the tariff, rather than to the tariff per se.

HARRY G. JOHNSON.

UNIVERSITY OF MANCHESTER

2. That is, by *B*'s marginal propensity to spend on imports, minus *A*'s marginal propensity to spend on home-produced goods (which is equal to one minus *A*'s marginal propensity to spend on imports).

3. Ozga's argument is expressed in terms of the effects of the combined changes in expenditure on excess demand rather than on the balance of payments, and he does not explain why the net result in the second case must be a deterioration.

REPLY

I do not think that it will serve any useful purpose to speak in this context of an effect of tariffs per se. Professor Johnson means by it the effect of tariffs accompanied by such a change in aggregate expenditure as is necessary to enable the buyers to purchase the same quantities of goods as before the imposition of the tariffs. But

this has no better claim to be called the effect of tariffs *per se* than would have the effect of tariffs accompanied by such a change in aggregate expenditure as is necessary to maintain internal balance (as in my article); or, indeed, by any other. Johnson's division of the ultimate change in the balance of payments into a positive part to be ascribed to the effect of tariffs and another part (positive or negative) to be ascribed to the effect of changes in aggregate expenditure is thus quite arbitrary. What he is doing is in fact like trying to determine how much of a bell's tolling is due to the swinging of the bell and how much to the swinging of its clapper.

The line I took in my article was that if the change in aggregate expenditure is induced by the imposition of tariffs — in accordance with some broad principles of economic policy, such as the maintenance of internal balance — then the whole change in the balance of payments is the effect of tariffs; because if tariffs were not imposed, there would not be any change either in aggregate expenditure or in the balance of payments. (If the bell did not swing, there would be neither movement of the clapper nor tolling.) Johnson has not convinced me that it is possible to say more than this.

S. A. OZGA.

LONDON SCHOOL OF ECONOMICS

RECENT PUBLICATIONS

- ALLEN, ROBERT LORING. Middle Eastern Economic Relations with the Soviet Union, Eastern Europe, and Mainland China. Charlottesville, Virginia: Department of Foreign Affairs, University of Virginia. 1958. pp. 128. \$1.00.
- ANGELOPOULOS, ANGELOS. Will the Atom Unite the World? London: The Bodley Head. 1957. pp. 262.
- AUSTRALIAN DEPARTMENT OF TRADE. British Manufacturers in Australia. Melbourne, Australia. 1957. pp. 47.
- . Survey of Manufacturing Activity in Australia. Melbourne, Australia. 1957. pp. 106.
- BALDWIN, K. D. S. The Niger Agricultural Project. Cambridge, Massachusetts: Harvard University Press. 1957. pp. 221. \$5.00.
- BOARMAN, P. M. Der Christ und die Soziale Marktwirtschaft. Stuttgart, Germany: W. Kohlhammer Verlag. 1955. pp. 130.
- BROWN, J. DOUGLAS AND HARBISON, FREDERICK. High-Talent Manpower For Science and Industry. Princeton, New Jersey: Industrial Relations Section, Princeton University. 1958. pp. 97. \$3.00.
- BUCKINGHAM, WALTER S. JR. Theoretical Economic Systems: A Comparative Analysis. New York: The Ronald Press Company. 1958. pp. 518. \$7.00.
- BUREAU OF THE CENSUS, U. S. GOVERNMENT. County and City Data Book, 1956. Washington, D. C.: U. S. Government Printing Office. 1957. pp. 565. \$4.50.
- CANADIAN DEPARTMENT OF LABOUR. Labour Organization in Canada — 1957 Edition. Ottawa, Canada: 1957. pp. 120.
- . Strikes and Lockouts in Canada. Ottawa, Canada. 1957. pp. 42. 35 cents.
- . Wage Rates and Hours of Labour. Ottawa, Canada. 1957. pp. 276. \$1.00.
- CHACKO, K. C. The Monetary and Fiscal Policy of India. Bombay, India: Vora and Company. 1957. pp. 386.
- CHAMBERLIN, EDWARD H. The Economic Analysis of Labor Union Power. Washington, D. C.: American Enterprise Association, Inc. 1958. pp. 48. \$1.00.
- CLEGG, H. A. AND ADAMS, REX. The Employers' Challenge. Oxford, England: Basil Blackwell. 1957. pp. 179. 21 shillings.
- DERBER, MILTON AND YOUNG, EDWIN. Labor and The New Deal. Madison, Wisconsin: The University of Wisconsin Press. 1958. pp. 393. \$6.00.
- DORFMAN, ROBERT; SAMUELSON, PAUL A.; AND SOLOW, ROBERT M. Linear Programming and Economic Analysis. New York: McGraw-Hill Book Company. 1958. pp. 527. \$10.00.
- ECONOMIST INTELLIGENCE UNIT, LTD. Britain and Europe — A Study of the Effects on British Manufacturing Industry of a Free Trade Area and the Common Market. London: The Shenval Press. 1957. pp. 297.
- FOUSEK, PETER G. Foreign Central Banking: The Instruments of Monetary Policy. New York: Federal Reserve Bank of New York. 1957. pp. 116.
- GOLDSMITH, RAYMOND W. Financial Intermediaries in the American Economy since 1900. Princeton, New Jersey: Princeton University Press. 1958. pp. 415. \$8.50.
- GORDON, LELAND JAMES. The Function of the Consumer In A Free Choice Economy. Westport, Connecticut: The Kazanjian Foundation Lectures. 1957. pp. 51.
- HARRIS, ABRAM L. Economics and Social Reform. New York: Harper and Brothers. 1958. pp. 357. \$5.00.

- HEALTH INFORMATION FOUNDATION. *An Inventory of Social and Economic Research in Health*. New York: 1957. pp. 327.
- HICKMAN, W. BRADDOCK. *Corporate Bonds: Quality and Investment Performance*. New York: National Bureau of Economic Research. 1957. pp. 33. 75 cents.
- JEANNENEY, J. M. AND PERROT, MARGUERITE. *Textes de Droit Economique et Social Francais*. Paris: Librairie Armand Colin. 1957. pp. 711.
- JEWKES, JOHN; SAWYERS, DAVID; AND STILLERMAN, RICHARD. *The Sources of Invention*. New York: St. Martin's Press, Inc. 1958. pp. 428. \$6.75.
- JONES, L. *Shipbuilding in Britain*. Cardiff, Glamorganshire, Wales: The University of Wales Press. 1957. pp. 244. 30 shillings.
- KERRISON, IRVINE L. H. (EDITOR). *Safe Management of Pension and Welfare Funds*. New Brunswick, New Jersey: Institute of Management and Labor Relations. Rutgers, The State University. 1957. pp. 54.
- LEISNER, H. H. *The Import Dependence of Britain and Western Germany: A Comparative Study*. Princeton, New Jersey: International Finance Section, Department of Economics and Sociology. 1957. pp. 84. 25 cents.
- LUZARDO, RODOLFO. *Venezuela Business and Finances*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc. 1957. pp. 167. \$4.95.
- MALISSEN, MARCEL. *Investissement et Financement*. Paris: Librairie Armand Colin. 1957. pp. 215.
- MIKESELL, RAYMOND F. *Agricultural Surpluses and Export Policy*. Washington, D. C.: American Enterprise Association. 1958. pp. 52. \$1.00.
- MUKERJI, KARUNAMOY. *The Problem of Land Transfer*. Calcutta, India: Visva-Bharati University. 1957. pp. 274. 10 Rupees.
- NATIONAL ACCOUNTS REVIEW COMMITTEE OF THE NATIONAL BUREAU OF ECONOMIC RESEARCH. *The National Economic Accounts of the United States*. Washington, D. C.: United States Government Printing Office. 1957. pp. 202. 65 cents.
- NATIONAL COUNCIL OF APPLIED ECONOMIC RESEARCH. *Foreign Exchange. Crisis and The Plan*. Bombay, India: Asia Publishing House. 1958. pp. 31.
- . *Taxation and Foreign Investment*. Bombay, India: Asia Publishing House. 1957. pp. 164. \$9.50.
- NEW ENGLAND GOVERNORS' COMMITTEE ON PUBLIC TRANSPORTATION. *Public Transportation for New England*. Boston, Massachusetts. 1957.
- OHKAWA, KAZUSHI. *The Growth Rate of The Japanese Economy Since 1878*. Kunitachi, Tokyo, Japan: The Institute of Economic Research. 1957. pp. 250.
- PERLMAN, MARK. *Labor Union Theories in America*. Evanston, Illinois: Row, Peterson and Company. 1958. pp. 313. \$6.00.
- POTTINGER, DAVID T. *The French Book Trade in the Ancient Regime, 1500-1791*. Cambridge, Massachusetts: Harvard University Press. 1958. pp. 363. \$7.50.
- POUNDS, NORMAN J. G. AND PARKER, WILLIAM N. *Coal and Steel in Western Europe*. Bloomington, Indiana: Indiana University Press. 1957. pp. 381. \$6.00.
- RESEARCH AND POLICY COMMITTEE. *Economic Growth in the United States: Its Past and Future*. New York: The Committee for Economic Development. 1958. pp. 63.
- ROBERTSON, DENNIS H. *Lectures on Economic Principles, Volume I*. New York: John de Graff, Inc. 1957. pp. 172. \$3.50.
- SASTRI, J. V. S. RAMA. *Nationalization and the Managerial Role*. Bombay, India: The Popular Book Depot. 1957. pp. 119.
- SAULNIER, R. J.; HALCROW, HAROLD G.; AND JACOBY, NEIL. *Federal Lending and Loan Insurance*. Princeton, New Jersey: Princeton University Press. 1958. pp. 566. \$12.00.

- SAYERS, R. S. *Modern Banking*, 4th Edition. New York: Oxford University Press. 1948. pp. 330. \$3.40.
- SHACKLE, G. L. S. *Time in Economics*. Amsterdam, Holland: North-Holland Publishing Company. 1958. pp. 111. \$6.50.
- SMITH, HOWARD R. *Government and Business — A Study in Economic Evolution*. New York: The Ronald Press Company. 1958. pp. 802. \$7.50.
- TIMOSHENKO, VLADIMIR AND SWERLING, BORIS. *The World's Sugar*. Stanford, California: Stanford University Press. 1957. pp. 364. \$6.50.
- U. S. JOINT ECONOMIC COMMITTEE. *Policy for Commercial Agriculture — Its Relation to Economic Growth and Stability*. Washington, D. C. 1957. pp. 864.
- VINER, JACOB. *Commercio Internazionale E Sviluppo Economico*. Torin, Italy: Unione Tipografico-Editrice Torinese. 1957. pp. 808.
- WADIA, P. A. AND MERCHANT, K. T. *Our Economic Problem*. Bombay, India: Vora and Company. 1957. pp. 825.
- WALKER, MABEL. *Business Enterprise and the City*. Princeton, New Jersey: Tax Institute, Inc. 1957. pp. 144. \$3.00.
- WEINTRAUB, SIDNEY. *An Approach to the Theory of Income Distribution*. Philadelphia, Pennsylvania: Chilton Company. 1958. pp. 214. \$6.50.
- WICKSELL, KNUT. *Selected Papers on Economic Theory*. Cambridge, Massachusetts: Harvard University Press. 1958. pp. 292. \$6.50.
- YAMANAKA, TOKUTARO AND KOBAYASHI, YOSHIO. *The History and Structure of Japan's Small and Medium Industries — With Two Specific Surveys*. Tokyo: Science Council of Japan, Division of Economics and Commerce. 1957. pp. 89.

American Economic Review

CONTENTS

VOLUME XLVIII

MARCH 1958

NUMBER 1

ARTICLES

- | | |
|---|---|
| Institutionalism and Welfare Economics | <i>M. A. Copeland</i> |
| Changes in Retailers' Buying | <i>R. P. Mack and Victor Zarnowitz</i> |
| The Supply of Storage | <i>M. J. Brennan</i> |
| Foreign Trade, Balanced-Budget Multipliers | <i>F. D. Holzman and Arnold Zellner</i> |
| The Clayton Act: Sleeping Giant of Antitrust? | <i>R. W. Harbeson</i> |

COMMUNICATIONS

- | | |
|-------------------------------------|------------------------------------|
| Time-Preference and Economic Growth | <i>Wassily Leontief</i> |
| Neutralization of Currency Flows | <i>L. M. Valentine</i> |
| Intermediaries and Monetary Theory | <i>J. M. Culbertson</i> |
| Reply | <i>J. G. Curley and E. S. Shaw</i> |

The AMERICAN ECONOMIC REVIEW, a quarterly, is the official publication of the American Economic Association and is sent to all members. The annual dues are \$6.00. Address editorial communications to Dr. Bernard F. Haley, Editor, American Economic Review, Stanford University, Room 220, Stanford, California; for information concerning other publications and activities of the Association, communicate with the Secretary-Treasurer, Dr. James Washington Bell, American Economic Association, Northwestern University, Evanston, Illinois. Send for information booklet.

ÉCONOMIE APPLIQUÉE

Tome XI

JANVIER-MARS, 1958

No. 1

STABILITÉ et PROGRES (Congrès de l'Association Economique internationale, Rome — 1956)

D. Robertson: Stabilité et Progrès: le problème des pays riches. *J. Viner*: Stabilité et Progrès: les problèmes de la pauvreté. *F. Perroux*: La recherche de la Stabilité — sur la différence entre les politiques anticycliques et les politiques de croissance harmonisée. *G. Haberler*: Facteurs monétaires et stabilité économique. *E. Lundberg*: Stabilité internationale et économies nationales. Discussions
Cahiers, *J. Lacroix*, *R. Ruyer*, *J. Y. Calvez*, *G. Granger*, *G. de Greef*: Recherches et dialogues philosophiques et économiques (n° 55). *J. R. Bouderville*: Pôles de croissance brésiliens — la sidérurgie du Minas Gerais (n° 56). *L. Couffignal*, *W. Ross Ashby*, *W. Grey Walter*, *R. Vallée*: Etudes sur la Cybernétique et l'Economie (n° 57). *G. Gruson*: La prévision économique aux Etats-Unis (n° 58). *R. Bertrand*: Analyse du tarif douanier français par industries (n° 61). *J. Weiller*, *J. Duvaux*: Economie française, échanges extérieurs et structures internationales (n° 62). *J. Le Lannou*, *J. Wolfe*: Brésil (n° 63). *F. Perroux*: Théorie générale du progrès économique — Les mesures des progrès économiques et l'idée d'économie progressive (n° 47); Les composants: la Création (n° 59) la Propagation, modèles microéconomiques (n° 60).

Economie Appliquée et Cahiers de l'Institut de Science Economique Appliquée

Directeur: François Perroux

Secrétaire Général: H. Bonjour

(I.S.E.A., 35 Boulevard des Capucines, Paris 2ème)

THE ECONOMIC JOURNAL

The Quarterly Journal of the Royal Economic Society

No. 269

MARCH 1958

Vol. LXVIII

I. Articles

Equilibrium and Disequilibrium: Misplaced Concreteness and Disguised Politics

F. Machlup

On Psychological Treatment for the Classics (Review Article)

A. L. Macfie

The Economic Framework of a Theory of Growth

J. H. Power

The Rate of Interest, the Marginal Efficiency of Capital and Investment Programming

W. H. White

The 1957 Investigation of the Financial Condition of the United States

S. E. Harris

The Problem of the Peak, with special reference to Road Passenger Transport

G. J. Ponsonby

Notes on Countervailing Power

A. Hunter

British Inter-Industrial Earnings Differentials, 1924-55

P. Haddy and M. E. Currell

II. Reviews

III. Notes and Memoranda

IV. Recent Periodicals and New Books

Macmillan and Co., Ltd., London, W.C. 2

Application for Fellowship to

Royal Economic Society, 21 Bentinck Street, London, W. 1

Annual Subscription £1 10s.

Single Copy Ten Shillings

THE JOURNAL OF BUSINESS

SCHOOL OF BUSINESS, UNIVERSITY OF CHICAGO, CHICAGO 37, ILLINOIS

Volume XXXI

APRIL 1958

Number 2

Business Objectives and Survival Needs: Notes on a Discipline of Business

Enterprise *Peter F. Drucker*

Transferable Skills and Abilities..... *Theodore Yntema*

Time, Demand, and Market Position..... *Yale Brozen*

Collective Bargaining and Management Functions: An Empirical Study

Milton Derber, W. Ellison Chalmers, Ross Stagner

Tax Reductions and Tax Deferrals..... *John W. Coughlan*

Businessmen as Regulatory Commissioners..... *Lincoln Smith*

Books Received

Notes: University Schools of Business

Book Reviews

The JOURNAL OF BUSINESS is published quarterly by the School of Business of the University of Chicago. Subscriptions are \$6.00 per year and should be addressed to the University of Chicago Press, Journals Division, 5750 Ellis Avenue, Chicago 37, Illinois. Editorial correspondence should be addressed to Irving Schweiger, Editor, JOURNAL OF BUSINESS, University of Chicago.

METROECONOMICA

INTERNATIONAL REVIEW OF ECONOMICS

Volume X Number 1

Editorial

METROECONOMICA is ten years old .

Business investment and marginal cost, *N. J. Cunningham*; Le principe de Pareto et la théorie de l'interet général, *R. Dehem*; Economic development and the balance of payments, *K. K. Kurihara*; An analogy between Marshall and Keynes, *C. E. Ferguson*; Reinvestment cycles and depreciation reserves under declining balance depreciation, *E. Schiff*; A macrodynamic model of growth, *A. Ll. Wright*.

Board of Editors: C. Bresciani Turrone, L. Dupriez, E. Fossati (*Managing Editor*), R. Frisch, W. Leontief, G. Lutfalla, A. Mahr, J. Marschak, G. U. Papi, R. Roy, E. Schneider, G. L. S. Shackle, R. Stone, J. Tinbergen, G. Tintner, F. Zeuthen.

Published every four months. Subscription price \$7.00 per year (two years \$12.00). Communications to the editor to be addressed to *Professor Eraldo Fossati*, director of the "Laboratorio di economia Vilfredo Pareto" University, Genova (Italy). Communications regarding subscriptions should be sent to *Licinio Cappelli*, Corso 12, Trieste (Italy) or *G. E. Stechert-Hafner Inc.*, 31 East 10th Street, New York, 3, N.Y. (U.S.A.)

REVUE DE SCIENCE FINANCIÈRE

THE FRENCH REVIEW OF PUBLIC FINANCE

Number II. April-June, 1958

The Budget Deficit and Its Financing

The Rational Use of Natural Resources and the Balance of Accounts

Tax Factors Affecting Foreign Investments in the United States

Tax Advantages in Respect of the Added Value Tax

From the Classical Budget Discussion to the Testing of the New Procedure

Survey of Foreign Finance

Survey of Jurisprudence

Book Reviews

Bibliography

French and International Financial Documents

R. Stucken (Erlangen)

A. Armengaud (Paris)

I. T. Wender (New York)

J. Monteil (Lyons)

J. Petot (Paris)

J. Lecaillon (Lille)

M. Chrétien (Lille)

The Review is published quarterly. Yearly Subscription abroad, 4,000 French francs or 9.50 U.S.A. dollars. Orders can be placed to Pichon and Durand-Auzias, 20 rue Soufflot, Paris (V°), and through Stechert-Hafner, 31 East 10th Street, New York, N. Y.

K Y K L O S

INTERNATIONAL REVIEW FOR SOCIAL SCIENCES

VOL. XI — 1958 — FASC. 1

Wage Determination Revisited

Oikonomia: An Inquiry into Beginnings of Economic Thought and Language

A Dynamic Interpretation of Malthus' Principle of Population

Theoretical Welfare Economics: A Comment

American Sociology for Europe?

An Analysis of Modern Inflation

J. Pen

Kurt Singer

John E. Moes

M. W. Reder

F. Zweig

Paolo Baffi

BOOK REVIEWS — TITLES OF NEW BOOKS

All articles are followed by a summary in English, French and German. Published quarterly. Annual subscription Sw.Fr. 26. — (U.S. \$6.00), postage included. Order from your bookstore or the publisher: Kyklos-Verlag, Postfach 20 365, Basel 2, Switzerland.

PUBLIC FINANCE / FINANCES PUBLIQUES

International quarterly journal devoted to the study of fiscal policy and related problems

Publisher and Editor in Chief: DR. J. A. MONOD DE FROIDEVILLE

Emmalaan 2, Haarlem, Netherlands

VOLUME XII

1957

NUMBER 4

G. Schmolders, Cologne Armament Price Control

Frank H. Jackson, Florida State University,

Tallahassee, U. S. A. Political Aspects of Export Taxation

B. R. Misra, Patna, India The Finance Commission

M. Masoin, Belgium Institute International de Finances Publiques

René Roux, Genève The Budgetary Reform of Technical Assistance

John F. Due, Urbana, Illinois The Literature of Sales Taxation

Contributions in French or German are followed by an English Summary.

Subscription Price for 1958: \$10.50, postage free.



USERS

EMPLOYEES

INVESTORS

The Three-way Benefits of Good Telephone Earnings

They benefit the telephone customer by providing the means to expand and improve the service and do it economically.

They benefit employees because they help to provide good jobs.

They benefit the investor by protecting his savings and insuring a good and secure return on his investment.

There is, indeed, no basis for the belief that keeping telephone earnings low is a sure road to keeping rates low.

Such a philosophy, by limiting progress and long-pull economies, will lead almost always to the opposite result . . . poorer service at a higher price than the customer would otherwise have to pay.

In all lines of business, it's the companies whose earnings are good that are able to make the best products, provide the best service, and give the best values.

BELL TELEPHONE SYSTEM



Timely New McGRAW-HILL Books

ECONOMICS: An Introductory Analysis

By PAUL A. SAMUELSON, Massachusetts Institute of Technology. New Fourth Edition. Ready in May

A new revision of one of the most competent and widely adopted works ever written for a beginning course. Utilizing an over-all national income approach, it synthesizes the basic principles of price and distribution. Greater emphasis is placed on inflation and monetary policy, and on the micro-economics of pricing. An important new chapter on "Economics of Atomic Energy and Automation" has been added. A Study Guide and Workbook will also be available in May. An up-to-date collection of "Readings In Economics" by P. A. Samuelson, R. L. Bishop, and J. R. Coleman, has just been published. (Third Edition: 474 pages, \$4.50, paper edition — \$3.25).

ECONOMIC FORECASTING

By V. LEWIS BASSIE, University of Illinois. Ready in September

Here is a new approach, combining dynamic economics and business forecasting and including more new material than any recent book in the field. It is the result of 10 years of research by the author. Part I reviews the various forecasting methods developed in the past; Part II shows in detail one method of working out a forecast of the over-all movements of the economy; Part III deals with important forecasting problems of such fields as public policy and business operations.

BUSINESS CYCLES AND ECONOMIC GROWTH

By JAMES S. DUESENBERY, Harvard University. *McGraw-Hill Economics Handbook Series*. 335 pages, \$6.50

This timely new economics text offers a thorough explanation of income movement and its components with a simplified description of economic behavior. The theory of income movement provides a basis for analysis of the impact of structural change. Chapters on "Marginal Investment Efficiency"; "Oligopoly and Investment"; and "Wages and Labor Supply" are included.

INTERNATIONAL AND INTERREGIONAL ECONOMICS

By SEYMOUR E. HARRIS, Harvard University. *McGraw-Hill Economics Handbook Series*. 564 pages, \$7.00

Concerned with the theory of international economics and its connection with the problems of the real world of 1946-1956. In particular, Professor Harris has stressed the disequilibrium of this period, and has summarized the history of trade, monetary and fiscal theory insofar as it applies to international economics. There is particular emphasis on such problems as dollar shortage, the European Payments Union, and the inadequacies of the free trade theory.

McGraw-Hill Book Company, Inc.

330 West 42nd Street

New York 36, N.Y.

Send for
copies on
approval

PRINCETON UNIVERSITY PRESS

Federal Lending and Loan Insurance

By R. J. Saulnier, Harold G. Halcrow & Neil H. Jacoby

Until now the lending and loan insuring activities of federal and federally sponsored agencies could be viewed only piecemeal, through the various agencies' reports. In this volume, three distinguished economists present a comprehensive review and analysis of government credit-granting and related activities in all sectors of the economy since 1917. Government credit programs are reviewed in the light of their historical development, present position, lending experience, and impact on private finance and the American economy.

Published for the National Bureau of Economic Research.

596 PAGES. \$12.00

Corporate Bond Quality and Investor Experience

By W. Braddock Hickman

This volume of the Corporate Bond Research Project provides an analytical summary of the Project's extensive body of tabulated material. It traces movements in the volume of high-grade and low-grade securities under different economic conditions, and examines the experience record of different classes of bonds. This book is closely related to *The Volume of Corporate Bond Financing since 1900* by the same author. *Published for the National Bureau of Economic Research.*

584 PAGES. TABLES AND CHARTS. \$10.00

The Income-Tax Burden on Stockholders

By Daniel M. Holland

In this study Dr. Holland is concerned with measuring the degree to which stockholders have been subject to heavier (or lighter) tax liabilities (because of corporation income tax and the personal tax on dividends), and how this differential burden varies with size of income. *Published for the National Bureau of Economic Research.*

256 PAGES. TABLES AND CHARTS. \$5.00

Order from your bookstore, or

Princeton University Press, Princeton, N. J.

MANAGEMENT PRINCIPLES AND PRACTICES

BY DALTON E. MC FARLAND, MICHIGAN STATE UNIVERSITY

An analytical survey of all phases of management, this text incorporates important new research in the social science and human relations field. Pertinent case material drawn from actual business situations makes this a practical, realistic text for introductory courses in management.

Coming Spring 1958

THE INTERNATIONAL ECONOMY, Revised Edition

BY PAUL T. ELLSWORTH, UNIVERSITY OF WISCONSIN

An up-to-date survey of international economics, this revised text covers both the theoretical principles determining international specialization, trade, and balance of payments adjustment, and the historical changes in the social and economic setting within which these principles have operated.

Coming Spring 1958

ESSENTIALS OF ACCOUNTING

BY WILLIAM A. PATON AND ROBERT L. DIXON, UNIVERSITY OF MICHIGAN

An exceptionally full treatment of accounting fundamentals, this successful text reflects recent thinking in accounting, with important chapters dealing with consolidated statements and the impact of inflation on accounting measurements. Highlighted throughout the text are the management point of view and the importance of cost accounting. *Problems and Practice Sets (plus Ruled Forms)* to accompany the text.

Coming Spring 1958

HOUSING: A Factual Analysis

BY GLENN H. BEYER, CORNELL UNIVERSITY

Discussing important considerations in home planning and design based on current research, this text covers demand and supply factors of the market; production, financing and tenure; and design criteria and environmental factors. Attractively illustrated by 72 photographs and drawings.

1958, 355 pages, \$6.75

The Macmillan Company

60 FIFTH AVENUE, NEW YORK 11

THE

QUARTERLY JOURNAL OF ECONOMICS

FOUNDED 1886

Vol. LXXII

August, 1958

No. 3

THE EXTREMITIES OF CURRENT AGRICULTURAL POLICY PROPOSALS.....	JOHN D. BLACK	307
THE RATE OF INTEREST IN A DYNAMIC MODEL	A. LL. WRIGHT	327
THE ANATOMY OF MARKET FAILURE.....	FRANCIS M. BATOR	351
A CRITIQUE OF CONCEPTS OF WORKABLE COMPETITION	STEPHEN H. SOSNICK	380
ECONOMIC STRUCTURE AND THE REGULATION OF TELEVISION	HARVEY J. LEVIN	424
NOTES AND DISCUSSIONS		
Peak Loads and Efficient Pricing:		
Comment.....	Jack Hirshleifer	451
Further Comment.....	H. S. Houthakker	463
Reply.....	Peter O. Steiner	465
Investment Criteria and Capital Intensity Once Again	Albert O. Hirschman and Gerald Sirkin	469
The Rockefeller Report: Implications for Fiscal and Monetary Policy	Arthur Smithies	472
Recent Publications.....		475

PUBLISHED BY HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS

\$1.50 PER COPY

\$5.00 A YEAR

The Quarterly Journal of Economics

Published by Harvard University

Editor

ARTHUR SMITHIES

Associate Editors

ABRAM BERGSON
JOHN D. BLACK*
J. K. BUTTERS
T. N. CARVER*
EDWARD H. CHAMBERLIN
ROBERT DORFMAN
JAMES S. DUESENBERY
JOHN T. DUNLOP
JOHN KENNETH GALBRAITH
ALEXANDER GERSCHENKRON
GOTTFRIED HABELER
* Emeritus

ALVIN H. HANSEN*
SEYMOUR E. HARRIS
CARL KAYSEN
WASSILY LEONTIEF
EDWARD S. MASON
GUY H. ORCUTT
SUMNER H. SLICHTER
OVERTON H. TAYLOR
ABBOT P. USHER*
JOHN H. WILLIAMS*
EDWIN B. WILSON*

Assistant Editor

CATHERINE RUGGLES GERRISH

Books, periodicals, and manuscripts to be addressed, THE QUARTERLY JOURNAL OF ECONOMICS, M-12, Littauer Center, Cambridge, Mass.

Business letters to be addressed, HARVARD UNIVERSITY PRESS, 79 Garden Street, Cambridge 38, Mass.

Subscriptions \$5.00 a year. Three years, \$14.00. Single copies (except those published in two parts) \$1.50, postpaid. A special subscription rate of \$3.00 is offered to students. Such special subscriptions may be for one, two, three, or four years. Those desiring to take advantage of this reduced rate should include with their remittance a statement from an officer of their college or university that they are *at that time* "regularly enrolled students who have not yet received the Doctor's degree."

CONTENTS FOR MAY, 1958

A LETTER FROM ADAM SMITH.....	Z. Clark Dickinson
IS THE INCOME OF THE "CADRES" A SPECIAL CLASS OF WAGES?.....	Jean Marchal and Jacques Lecalllon
INTERNATIONAL SPECIALIZATION AND THE CONCEPT OF BALANCED GROWTH.....	John Sheahan
WHAT IS THE ECONOMIC SYSTEM?.....	David McCord Wright
THE CLASSICAL ECONOMISTS AND THE FACTORY ACTS — A RE-EXAMINATION.....	Mark Blaug
ADAPTIVE EXPECTATIONS AND COBWEB PHENOMENA.....	Marc Nerlove
COMPETITION, CREDIT POLICIES, AND THE CAPTIVE FINANCE COMPANY.....	Paul H. Banner
LOCATION AND THE THEORY OF PRODUCTION.....	Leon N. Moses
THE EFFECT OF DEVALUATION ON THE PRICE LEVEL.....	Egon Sohmen
REVIEWS, NOTES AND DISCUSSIONS.....	
The Business Cycle In the Post-War World: A Review.....	James Tobin
Investments in United States Government Securities by Nonfinancial Corporations, 1952-1956:	
Comment.....	Colin D. Campbell
Reply.....	Hyman P. Minsky
Tariffs and the Balance of Payments:	
Comment.....	Harry G. Johnson
Reply.....	S. A. Ozga
Recent Publications.....	

COPYRIGHT, 1958, BY THE PRESIDENT AND FELLOWS OF HARVARD COLLEGE

Entered as Second-Class Mail Matter. Acceptance for mailing at special postage rate provided for by Section 1103, Act of October 3, 1917, authorized on July 31, 1918
Published by the Harvard University Press, Cambridge, Mass., U.S.A.
London: Oxford University Press

THE QUARTERLY JOURNAL OF ECONOMICS

Vol. LXXII

August, 1958

No. 3

THE EXTREMITIES OF CURRENT AGRICULTURAL POLICY PROPOSALS

By JOHN D. BLACK*

I. Various measures proposed, 307. — II. A more moderate program, 315.
— III. Conclusion, 323.

I

The general situation now with respect to farm policy in this country is that moderate, intermediate and often compromise measures having failed to solve the farm problem, different groups are now proposing extreme measures. These extremes, however, are not all in the same direction. They range from practically "free markets" to the reverse of these; and also from what is largely government control to producer-group control.

Take first what must be considered as one of the extremes, the policy advocated by President Shuman of the American Farm Bureau Federation at its recent annual convention. This includes abolishing at as fast a rate as possible all price-supports except as "a safeguard against extremely wide fluctuations" and likewise all acreage allotments and marketing quotas, and stopping all flow of commodities into Commodity Credit Corporation stockpiles and disposing of present stocks though Public Law 480 as fast as possible, except as loans and purchases may be needed in emergency low-price situations. Even in such situations, the stocks are to be disposed of promptly at home or abroad. There is, of course, the definite assumption in all this that if the government really gets moving in this direction, all these actions will be "possible," and that the present tangled situation will straighten itself out, and that thereafter only in extreme situations will the government need to intervene again. As we shall see later, the official pronouncement of the Farm Bureau Federation at the end of the convention did not go to the extreme of President Shuman's statement. It did call for a return to essentially free-market operation both domestically and internationally and as

* Harvard University.

"rapidly as possible," but it allowed for gradualness in the movement to this end.

At the other extreme from Shuman's statement as to government participation is the demand for putting not only all the present so-called "basic" commodities — cotton, wheat, tobacco, corn, peanuts, and rice — but the remaining feed crops as well on a strict individual farm marketing quota basis with price supports at 90 per cent of parity or better. The requirement of a referendum is still included in such proposals, it being assumed that the vote would be strongly for such quotas given such a level of prices, as it has been in the case of tobacco. The method of price support commonly specified is that of loans and purchases, those not keeping within their quotas to be largely denied such loans and purchases.

A detail of this proposal specified by some is that the present exemption of small acreages of these crops — such as the fifteen-acre exemption for wheat — be largely removed. The wheat now being produced under this exemption on farms in the Corn Belt and East and South is estimated to have reached close to 100 million bushels in recent years.

All of the "basic" crops except corn are already on a potential over-all marketing quota basis, but the individual farm quotas are stated on an acreage basis. The reason for the marketing quota provisions is that the farmers by using more fertilizer or other improved technology on their restricted acreages have increased their yields per acre to a level that largely offsets the fewer acres. Thus the 15,560,000 acres of cotton in 1956 averaged 408 pounds per acre compared with the 239 pounds on 27,000,000 acres in 1936-40. The corn yields have increased from 26 to 54 bushels per acre in the same period; the tobacco yields from 910 to 1570 pounds. Most of these increases have come since 1946-50; thus 50 of the 70 per cent for cotton, 150 of the 210 for corn, but only 30 of the 70 for tobacco.

Particularly to be noted in this connection is that few of the extreme suggestions go so far as to put livestock numbers or sales on a quota basis. The proposals are to control livestock output only by controlling the feed supply. The main reason for this is that direct control of livestock output is generally believed to be unworkable. With all of the "basic" crops except corn already on an over-all market quota basis, it is the extension of such quotas to corn and the nonbasic feed crops that is most important. Large acreages of land that normally grow wheat or cotton can be shifted very simply to corn, grain sorghum or oats and other feed grains.

The present marketing quotas on the basic crops are also limited

in their effectiveness because the marketing quotas are converted to acreage quotas by dividing them by the prevailing average yields and then lower limits are set on the total acreage quotas — thus 55 million acres for wheat and 17.4 million for cotton. Without these limitations and with present stocks of wheat and cotton, these acreages could be cut more than a half under other provisions of the laws now in force. Proposals are being made, of course, for greater latitude in these acreage cuts, but still most of the producers of the basic crops must be left with some productive use to which they can put their land.

The most logically consistent of all the proposals along the foregoing lines is that of Professor Willard Cochrane of the University of Minnesota, presented to the Joint Economic Committee of Congress last November. Cochrane would have agriculture in effect declared a public utility like the railroads and would have the government each year set a national sales quota for each important farm product and a price that would bring production and consumption into balance at that price. This price, however, would be “fair” to both producers and consumers, a “fair” price to producers being defined as one equal to “parity” and yielding as good an income to farmers as that received by other groups in the nation. The government would determine the level of such parity. Clearly enough, these prices would be monopoly prices enforced by the government by keeping supplies at a level that would support such prices in the market place. The national sales quota would be broken down into individual producer sales quotas, nation-wide percentage departures from these being announced year by year according to changing export demand, carry-over stocks, and the like. Also the sales quota certificates would be transferable. Cochrane would limit this monopoly control to fifteen to twenty-five products, and exclude feed grains, baby chicks and other products being sold mostly to other farmers. The nearest antecedent to such a program, Cochrane finds in the present sugar control program. But he points out that in effect the program for tobacco is not too far removed, and that both of these programs have become very acceptable to the producers of these products.

The program put forward before this same Joint Economic Committee by Glen Talbott, Chairman of the Executive Committee of the National Farmers' Union, the nation's most nearly leftist farm organization, differs from Cochrane's mainly in the following particulars:

1. A government board “composed mainly of democratically elected farmers” will set the level of prices and over-all national sales

quotas, designated as the "all-commodity farm marketing goal."

2. The level of prices is to assure the farmers "parity income" instead of "fair" prices.

3. A national "voluntary conservation acreage reserve" goal is to be set, which will include all the land not needed to produce the "all-commodity farm marketing goal," the farmers will receive offers of annual rentals from the Secretary of Agriculture to keep this acreage reserve out of commercial production, plus payments to cover the costs of conservation practices.

4. The farmer will have the option, however, of using land in excess of the "all-commodity goal certificate" for his farm in production at 75 per cent of the parity income rate.

5. Government credit is to be made available to enable farmers co-operatively to take over as fast as they are able the whole business of assembling, processing, storing and marketing farm products (as now done in a large way by the Farmers' Terminal Grain Marketing Association in the northern Great Plains).

6. This is to be an "all commodity" program, but provision is made for separate programs of two types:

(a) Those patterned after the federal orders now used in city fluid milk markets and for some vegetables, fruits, and nuts.

(b) Separate commodity goal programs for some commodities not protected by the general program.

The reasoning basic to the extreme proposals of the Cochrane type is simply that the farmers left to themselves have little bargaining power in the market, nor will they organize into bargaining unions like labor; therefore the government must step in and function largely in lieu of the market. The Talbott approach calls for a large measure of participation by farmers and their representatives in the setting of marketing goals or quotas, and in marketing operations and pricing in the market place. It also leaves more choice to the producer. It therefore departs considerably from a straight public utility program.

In fact, the Farmers' Union proposal has in it much of the philosophy of the Aaron Sapiro co-operative marketing movement which began in the early 1920's, and which ultimately found partial expression in the Hoover Agricultural Marketing Act of 1928 and the setting up of the Federal Farm Board. Sapiro designed his commodity co-operatives with a view to signing up 85 or 90 per cent of the production and feeding the supply to the market in an orderly manner and forcing the processors to pay full market values. The nationwide co-operatives that the Farm Board set up in 1924-30—grain, cotton, wool, beans, etc.—were partly amalgamations of Sapiro-

type co-operatives. Their undertakings to hold supplies out of the market as prices fell, of course, turned out badly because of the continuing sag of prices in the Great Depression of the 1930's, but under more normal circumstances these co-operatives might have survived for several years at least. Their operations, with government corporations like the grain and cotton stabilization corporations of 1930 to back them up, would not have been greatly different from those of the Commodity Credit Corporation in more recent times.

In the same general class as the Cochrane and Talbott proposals, and based on the same doctrine of monopoly power for agriculture to match that of business and labor, is a recent proposal, called a "self-help" program, that a tax of 10 to 15 per cent be collected at point of first sale on *all* farm products sold, and that the proceeds be used to buy up and dispose of, outside of domestic marketing channels, enough of each year's farm output, estimated at 5 per cent, to raise farm prices 30 to 40 per cent, and to buy or rent and hold out of use enough whole farms over the next five or ten years so that by the end of that time there will be no surpluses to buy. The "self-help" idea was first advanced in a program offered by the National Grange and the National Milk Producers' Federation, to apply to dairy products only, with the tax proceeds to be used only for buying up and disposing of the surplus. Both proposals call for setting up a board or "authority" representing the farm producers that will administer the program.

One statement alone should suffice at this point for the over-all "self-help" proposal, namely, with farm prices actually supported at 30 to 40 per cent above recent levels, yields per acre will rise so fast on the remaining farms, and also rents and land prices, that the whole program would break down. Even the Talbott proposal of 75 per cent of his income parity on the output of land not covered by goal certificates would suffer the same fate, since the additional cost of farming additional acres is lower even than present market prices.

The foregoing brings out another issue between the groups, as to how much is left to administrative discretion. The Secretary wants a large measure of control in the Department of Agriculture. At the other extreme are, first, those in Congress who would spell out very explicit directions in legislation; second, those who would set up strong nonpartisan and nonspecial-interest boards to make the decisions; third, those who would have farmer boards do this, as proposed by Talbott of the Farmers' Union; and fourth, those who propose a joint decision between the farmer boards and the government, as in the self-help proposals.

Any listing of extreme proposals must surely also include the Brannan Plan as offered by Brannan himself in 1949. It may well be that a majority of our citizens consider as extreme any proposal to make up in direct payments to producers the difference between the market prices and some level of prices deemed more reasonable. But surely to do this at the level defined by Brannan, and without much regard to the size of the crop, was indeed extreme. It meant in effect saying to each producer after he had received this high price for a big crop, "Go ahead and produce another big crop. We guarantee you the same price for it." It would be difficult to conceive of any better way of inducing growing surpluses.

Certainly there are wide differences in the level of prices proposed to be supported, from a bottom of 60 per cent in recent administration proposals, to the 100 per cent demanded by the Farmers' Union and some other groups. The level most commonly named in the recent commodity-by-commodity proposals — wheat, cotton, dairy products, for example — is 90 per cent of parity.

Something needs to be said at this point about varying the level of price supports with the size of the supply. Provision for such variation was written into the Agricultural Acts of 1948 and 1949 under the term "flexible." The schedules now in use vary the support level for corn, wheat, and rice from 90 down to 75 per cent of parity as a bottom limit as the supply — current production plus carry-over — rises from 102 to 130 per cent of "normal"; for cotton and peanuts as the supply increases from 108 to 130 per cent or more of "normal." (These schedules do not apply to tobacco if marketing quotas are in effect.) The administration wants in effect to extend these schedules down to 60 per cent of normal with supplies over 130 per cent, and Secretary Benson at times has even suggested zero as the lower limit.

The most recent administration and Farm Bureau proposals, however, are to abandon these schedules and set only the lower limits. The administration is saying that raising the price-support level as supplies decline causes output to expand again. This raises the whole question of the slope of the supply-price curve in the market. The schedules now in use do not let the prices of these products fall as fast as they would in an open market. It has been their intent not to do so, and in this way to protect the farmers. If this is true, then prices will not rise as fast on these schedules with falling supplies as they would in the open market. Still the support price at these somewhat less ample supplies may be higher than it would be in an open market. What the administration apparently wants is freedom

to raise the price so little that it will not induce more output, and the Farm Bureau agrees with this view.

Then there is the difference between using as a support basis old parity — relative commodity-by-commodity prices mostly back in 1910-14 — or “modernized” parity, that is, relative moving-acreage prices in the last ten years. The extreme with respect to this standard is the level of whichever is the higher that was used in the early 1950’s. The level now in force is shifting from old to modernized parity at the rate of 5 per cent a year, reaching the latter for most products by 1960.

There are wide divergencies as to the definition of “income parity.” Those first proposing it, back in 1934-38, as a substitute for price parity had in mind simply taking account of the size of the crop as well as the price, so that farmers would not have the incentive to produce successively larger crops year after year. It was written in these terms into the Agricultural Act of 1938 as a basis for distributing the supplementary payments of that Act. But today income parity is being given by the Farmers’ Union and some other groups the extreme definition of the same returns per capita in farming as in other occupations. This would not be so extreme, of course, if there were official data available upon which to make a valid comparison of such returns. As I have shown on other occasions, those released by the U.S. Department of Agriculture showing about a 1-to-2 ratio are very far from valid, and if used as a basis for price or income supports would contribute importantly to perpetuating the low-income farming of our South, our cut-over regions, and other overpopulated areas.

This suggests another important divergence in the thinking of different groups in this country, between those who, at one extreme, want no further reduction in the number of our farmers and, at the other, want our farming to yield the highest returns possible per worker almost without regard to how few are the workers and farms and how farming is organized. Even the prospect of increasing integration of farming and business does not disturb them.

Related to the foregoing is the extreme position of the Farmers’ Union group that the surpluses that we have now, or would get with 100 per cent of parity prices, can be disposed of readily in either of two ways, or at least in the two combined. The first way is by helping low-income families in this country, by means of a food stamp plan or similar device, to include more meat, eggs and dairy products in their diets. They point out that a billion food calories in this form require six or seven times as much land to produce as a billion in the

form of cereals, potatoes, and sugar. The other way is to make our foods and cotton more readily available to ill-fed and ill-clothed people in other lands.

All that I shall say relative to this position is that more can advantageously be done in both these ways than is now being done, but any loose indiscriminating undertakings along these lines will be likely to hurt more than help. Merely keeping more hungry people alive year after year in an overpopulated country does not really help. Our supplies need to be used in ways that help them better provide for themselves thereafter. Working out joint undertakings to this end with these countries is a slow process.

Very different from the Farmers' Union on disposal of farm surpluses abroad is that of the commodity groups that are pressing for more of what has been called "international dumping" by its opponents. The State Department has in general struggled for discretion in such action, but not always successfully.

Better intentioned surely are those who would have the United States take the lead in setting up an internationally managed food reserve. But such a reserve must be handled in such a way that the surplus foods of any country move directly to the receiving countries without much delay. The costs of actual accumulation, storage, and subsequent distribution of such food supplies could easily equal twice the value of the food.

Other subjects in which there is a wide range of position, partly involved in the foregoing and partly not, are the following:

1. How large a stockpile of wheat, corn, cotton, and other farm products needs to be carried for national security and stabilization.
2. How large a conservation reserve of soil resources needs to be carried.
3. To what extent support to agriculture will be provided in the market and to what extent in direct payments.
4. How wide a list of products receive support.
5. What to do about low-income farming.

Surely what has been said demonstrates the opening statement as to the extreme measures that different groups are now proposing. The positions taken by these groups on many of the subjects are almost at opposite poles. Moreover, they have been becoming more so over the years since 1930, although some will be disposed to dispute this by saying that nothing as drastic as Wallace's pig slaughter has been undertaken or even suggested. This, however, was an exception. So far as price supports are concerned, all that was projected in the Act of 1933 was a gradual approach to parity.

II

From the standpoint of solving the farm problem, this growing divergence of position is indeed unfortunate, since it makes more difficult any agreement on a sane and effective course of action. Surely, moreover, it does not indicate a growth of the people's understanding of the problem, but rather the opposite — something a little bit in the nature of the divergence of opinion on the slavery issue that finally brought on the war between the states.

Let us therefore calmly ask ourselves if there is not some program of action between these extremes that offers promise of year-by-year improvement in the farm situation.

There can be no doubt that underlying much of this wide range of difference in position is a very elementary economic relationship, namely, that of the response of farm output to price. It was something like eight years ago that Secretary Charles Brannan said, in an auditorium lecture in the Department of Agriculture, very positively and definitely that there is no longer any supply-response relationship in agriculture. He virtually repeated this statement at Michigan State University last February. But if Brannan were right, there would have been no hog-corn cycles in the last eight years, nor any cattle cycle movement, nor the almost year-by-year ups and downs in poultry output. Moreover, it takes a down response as well as an up response to produce these cycle movements.

There is, however, a big difference between the response of the output of an individual product to price and that of groups of competing products, or of the total farm output of the nation. Part of the shrinkage in pork output usually is offset by an expansion in beef output; and similarly for wheat and grain sorghum, and grain sorghum and cotton. There has been, however, an over-all increase in whole groups of products, and in national farm output since 1952 in spite of lower prices. This is due to the strong upward trend in yields per acre and per animal unit and output per worker that set in during the war and has continued since. Few economists will not agree that the increase in output would have been considerably more if prices had stayed up at the postwar and Korean War levels, and that with no price supports there would have been a lesser increase, and possibly even a decline.

The reason for the increase in output with such price declines as has occurred has been, of course, that on very many of the larger or better farms producing probably two-thirds or more of the output, larger outputs yielded larger net cash incomes than would have

smaller ones, especially if more fertilizer and other improved technology was used. This has been especially true of farms with machines and equipment that were far from being fully used, as was commonly the case on what used to be family-sized farms.

What do such circumstances call for in the way of a public farm policy? What would the Shuman proposal of dropping all price supports as fast as possible achieve? Various agricultural economists have estimated that if all controls and supports were dropped and all supplies were dumped on the market, prices of farm products would drop 25 to 40 per cent further. This would cause disaster to several hundred thousand of our farmers with heavy indebtedness and serious hardship for most of the rest. It would today usually be something more serious on commercial farms than unemployment on social security.

Of course the Farm Bureau Federation itself is not proposing any such extreme policy. Its proposal is to move in this direction only as fast as possible. But what is "as fast as possible?" Does it mean "as fast as possible" *economically* or *politically*? The Farm Bureau proposals have now been spelled out more definitely in Congress as support of the farm prices of cotton and corn at a level of 90 per cent of the average of the last three years, and of other feed grains on the basis of their feed value relative to corn. Also a proposal to support other farm commodities now receiving price supports at this same 90 per cent level is being submitted to the State Farm Bureaus for their approval.

So far as corn is concerned, this 90 per cent three-year average will actually be a little higher than prices in the past year. The cotton three-year average is probably little if any lower than a large group of cotton industry leaders are now ready to go in order to save their home market from synthetics and their export market. But how far in this direction other commodity groups and even the rank and file of cotton growers, are ready to go is surely open to question. The Congressional vote of the South in March in favor of freezing prices and production quotas at 1957 levels was 95 for and 14 against; that of the Midwest 59 to 60; that of the West 24 to 26; and that of the Northeast, 31 to 73.

Granted political possibility, there is still the economic question whether this level of prices will be low enough to bring into balance the supplies and demand for the different products now receiving price supports. The probable best answer is, for some of them, yes, but only after time enough for the farmers to make their adjustments in production, and for the Commodity Credit Corporation to dispose

of whatever overstocks it has. For others of them, further reductions of the price-support level will be needed in later years, these probably best taking the form of step reductions, a few per cent at a time.

What, if anything, is wrong with such a program? It will really give us a solution in the end. One part of the answer is the time it will take, which will range from maybe as little as two or three years for a few products to as long as fifteen or twenty years for a few others. The reason for the longer periods for some commodities is that as indicated earlier, a large fraction of the more capable producers will meet the lower or lowering prices by more use of improved technologies that will lower their costs by increases in output per acre, per animal unit, and per man-hour. Of course, the size of the present overstocks is also a factor in the time it will take.

A larger part of the answer is the disaster or hardships that will be caused to a million or so of farm families who are not in a position or able by themselves to meet in good part the lowering prices with lowering unit costs, during these adjustment years.

A consequence of the foregoing is that solving this problem by nothing more than the measures proposed by the Farm Bureau Federation is not really politically possible; nor is it really economically advisable to subject such a large number of our farm families to such disasters and hardships.

I must also insist that neither is it equitable and fair for the nation to let these farm families suffer in this way. It was the greatly increased need for farm products by our allies in World War II and the Marshall Plan years following that mainly caused the great surge in the use of advanced technology that is giving us our present surpluses. It surely is not fair to agriculture to let it bear all the burden of the needed post-recovery adjustment. To be sure the nation is now spending vast sums on its agriculture sector; but it is not spending it in such a way as to achieve the needed adjustments. We must decide what these are, and devote our expenditures on agriculture to achieving these.

The Farm Bureau is right that a lower level of prices is needed. But this alone is far from enough.

First in order of the other things needed, I would put adaptations of our farm credit facilities to meet the needs of the situation that faces us. The first adaptation is ample protection against the foreclosure of mortgages that may threaten many farms at the lower product prices while readjustments are being made. Second is credit to buy the additional land that a million or so of our farms need in order to be economic units. Third is credit to buy the livestock

equipment and other things needed on these enlarged farms. Fourth is credit for these things and even for land improvements on some farms that can thus be made into economic units, this in spite of the general surplus situation. Even our existing public or public-co-operative credit agencies are now operating upon too restricted and conservative a basis to meet the needs of this situation. It is doubtful whether they can be changed from this practice and policy without some public action along three lines: *A.* Government underwriting or guaranteeing of large numbers of loans based upon approved budgeted farm and financial plans. These plans should spell out a time schedule of loans and adjustments. *B.* Building up an adequate and competent staff to work out these budgeted plans and supervise the loans subsequently. *C.* The government's giving definite orders to these credit agencies to make such loans. However, private credit agencies should be given a first chance to make any of these guaranteed loans. Loans of the foregoing types should be made to landlords as well as to owner-operating farmers provided the leasing arrangements are approved. Loans of the third type should be available also for tenants.

As an adjunct of this program, the Farm and Home Development Program activities should for the present be largely if not altogether absorbed in the planning of the farms being considered for such approved loans. Also all soil conservation districts receiving federal aid should be required to co-operate with the staffs making the farm and financial plans to the extent of making basic soil and land-use capability maps and conservation practice maps for these farms.

No provision is made in any of the foregoing for farm enlargement loans based upon additional rented land. This is intentional. Consideration might be given, however, to certain types of loans based on long-term leases.

With prices dropping around 10 per cent for most products all at once, however, the farmers obtaining such adjustment loans will be in difficulty making ends meet while the adjustments are under way. It will therefore be necessary to make supplementary income payments or the equivalent of these, to them for a few years.¹ The level here suggested for this is the difference between the proposed

1. The Farm Bureau is vigorously opposed to supplementary income payments. An equivalent of these more acceptable to it would be lowering the level of support prices by small steps over the same period of years. The supplementary income payments are preferred in this statement because with them the level of domestic open-market prices is the same as the export level, except in the case of import duties, as with wool, and both levels indicate almost from the start the price that brings supply and demand into balance.

announced loan support level and 90 per cent of modernized parity for the first year of the loan and thereafter declining regularly to one-half of this in the fifth year and at the same rate thereafter until a level of one-fourth is reached, after which there will be no payments.

Consideration should also be given to basing the income supplements not simply on a given amount per unit of product, but on the multiplication product of size of the crop and amount per unit, to avoid the evil effect of Brannan's plan. This would amount in effect to preserving something in the nature of the present schedule of price-support levels, but using size of the annual crop as a basis rather than supply.

What about the remainder of the farmers — those not needing additional loan facilities, those not able to qualify for such loans, etc. The range of circumstances among these is very wide. Consider first the large group of commercial farmers who have been meeting the situation pretty well thus far, so far as it concerns themselves, by means of improved technology, lower unit costs, etc. Even for these a 10 per cent drop in prices all at once is too severe to be reasonable or to be good economics. For them a level and year-by-year schedule of supplementary income payments is here suggested that is one-half that for the farmers adopting an adjustment loan program. The main reason for this distinction between the two groups is that most of these farmers are really in a position to make the needed adjustments by themselves. Another is the need for some incentive to induce many farmers to enter into adjustment loan programs for their farms. After all, it is this adjustment that is the prime objective of any program to solve the current agricultural problem.

Advisable, however, would be the making of the same level of income supplements available to farmers not needing loans, but entering into a contract to carry out the same kind of an adjustment program for their farms, as to those with loans in the agreement. Such provisions, both with and without loans, would take the place of the present Soil Bank program, at least the acreage reserve part of it and would put it on a much sounder and more constructive basis.

It is here assumed that the Agricultural Conservation Payments will be continued, except for fertilizer payments not fitting into the loan or adjustment plans. There still will be need for conservation reserve payments.

Something more specific needs to be said at this point as to the character and direction of the adjustments written into the loan or adjustment agreements. Making inadequate farms over into economic units is only part of the transition. With more acres of land

per farm, more of it can be in grass and pasture and in crops using more machine and less hand labor. This will have the effect of reducing the acreage output per farm acre though not per farm. Net incomes will be increased because of labor inputs and expenditures will be reduced more than sales receipts. The farm planning staffs working out the adjustment loan plans need to be thoroughly drilled in this aspect of their planning work.

As for the farm families at the other end of the scale, those not able to qualify for adjustment loans, any that are not too old should be helped in selling their farms and obtaining other employment, mostly nonfarm. In the latter category should be included many cropper and other tenant families on small units in the South and elsewhere. Some of both these groups, but especially of the latter, are going to need some financing to cover moving expenses and the like as well as help in finding a job. Definite procedures should be developed under which the local offices of the Federal Employment Agency and of the Farmers Home Administration work together in locating employment and planning needed loans for families and individuals registering for such help, the Farmers Home Administration making the loans and the Employment Agency helping with the supervision of them. This may seem to some like an undue extension of federal government humanitarianism, but surely it is not as extreme as a recent proposal for an outright grant of \$5000 to farm families ready to move into urban employment.

As for part-time farmers, some combination of the foregoing can be made to fit some of them. Of the farm families not included in the foregoing groups, many are too old for such vigorous measures as any of the foregoing, or too indisposed to change. Surely it is in the American tradition to grant them freedom to ride out their years as they want. Old Age Security is now available for them in time. The nation's main concern is that the young people in their families have a chance to develop.

An important modification of the foregoing is suggested in the case of export types of farm products, like cotton. It is that for all those for which net exports exceed some level — say 5 per cent of total production — the price support level would be no higher than the world trade level as reflected in domestic export markets, and that the supplementary income payments be fractions of the difference between this and 90 per cent of modernized parity in case this level is lower than the 90 per cent of the last three-year average proposed by the Farm Bureau.

The reason for this suggestion is that if supplementary income

payments are going to be made anyway, no more than simply making them a little larger is all that is needed to achieve a free interplay of the domestic and export markets, and such free interplay is desirable from the standpoint of international trade relations.

This is probably the best point at which to consider the feasibility of the domestic allotment — now called “domestic parity” — proposals. They all set too high a support level for the domestic quota — a level which combined with a lower price level on exports will give a composite price that will more than maintain present levels of over-production. They also make no direct attack on production adjustment as does the program outlined in the foregoing. They could, however, be revised in these respects. They have the advantage that the producers get their income supplements paid to them in the market place and hence involve no drains on the public treasury. Also in case producers are already on an allotment basis, it is relatively simple to split this between the domestic and surplus parts. It may well be that putting some farm products on this basis, with a support level in line with adjustment requirements, and lowering it further as may be needed in the future, would be preferable to handling all commodities the same way. It is here suggested that this be tried out for cotton, with revision in the two respects as just indicated.

Consideration is also needed of the production allotment feature of present and proposed programs. Even though they have proved generally inadequate in their present form, they have helped, and therefore should be continued in pretty much their present form until the supplies of any product are brought in line with demand by the other measures herein suggested, at which time they can cautiously be removed. It would not be wise at this time, however, to apply quotas to new products.

The proposals to impose stricter controls in the form of specified quantities to each individual producer must be considered from an administrative point of view. It would be difficult to determine such quotas that would be fair as between farms in the first place, and difficult to enforce them in the second. If they work as they would be supposed to, any farmer that increased his yields per acre more than others would have his acreage quota reduced. This could be most nearly achieved with crops like cotton and tobacco.

Next in order is the matter of how wide a list of farm products to include in a program. The Farm Bureau scheme of supports at 90 per cent of average prices of the preceding three years, even with the drastic modifications herein suggested, plus the adjustment loans and

all the rest, could conceivably be applied to any farm product. It would be advisable, however, to apply it at the outset only to those receiving the loan type of price supports in recent years, that is, the mandatory list, including the six basics plus dairy products, wool, honey, and possibly a few others.

What about the feed crops other than corn? They and corn are in heavy surplus now. They received some support in 1956-57. One way of handling these is to provide for their final products, namely, pork, beef, and dairy products, the kind of support herein suggested for the mandatory list. The average price used as a basis, however, should be that of the last hog-corn, cattle or other cycle, with adjustments for price trends. This would mean in effect the providing of income supplements to hog, beef, and dairy farmers at the low-price sectors in the cycles for these products. This has often been suggested as a way of leveling out these cycles. This would need to be very carefully done, however. It is doubtful if a forward-pricing plan for these products should name an average cycle price higher than 90 per cent of that which would develop without the forward pricing. An assured price brings forth more product than one that happens to average out at that level.

A near-final matter is the means by which prices are supported. Nothing different from the present Commodity Credit Corporation operations is suggested except the following:

1. Recognition of, and provision for, the nation's carrying-over larger stocks of a number of farm products than a free market will carry.

2. A significantly expanded program of distribution of surplus farm products among low-income families in this country. The best indication of what this could accomplish is to be found in a recent newsletter of the University of Minnesota.² The conclusions reached are that if all families in the United States with incomes of less than \$3,000 were given the food-buying power of the group with income of \$3,000 to \$4,000, the consumption of dairy products in this country would be raised by 12 per cent and that of meat, poultry, fish and eggs, 7 per cent, whereas that of flour and cereals would be decreased 19 per cent. The last of these would take away the market for the product of about 12 million acres of land now in wheat and other food cereals. The first two of these would require that around 180 million acres more land, or the equivalent of this, be devoted to pasture, forage and feed grains. The present proportions of land being used

2. *Minnesota Farm Business Notes*, November 27, 1957, John Wetmore and Willard Cochrane.

to feed livestock are about 125 millions in feed grains and hay, and 775 millions in pasture, most of this, of course, being semi-arid. For the part of the increases that would come from feed grains and hay, this would mean an increase of 25 million acres in these. This would in effect provide a use for a net of 13 million acres of our crop land now in surplus production.³ If some of the pasture increase came out of crop land, the net would be much larger. Workable procedures for bringing about such a shift in consumption are, however, not simple to devise, and will take time to produce their full effects.

3. Revision over the next few years of present procedures for distribution of agricultural surpluses abroad along the more constructive lines indicated earlier.

A final sentence or two is needed on the question raised earlier as to administrative decision-making in such a program. Surely no board of agricultural producers should perform in anything more than an advisory role. Also there is serious question whether the Secretary of Agriculture and his staff can be independent enough politically to make the best decisions. This leaves us with the third alternative, having a strong nonpartisan and nonspecial-interest board set up to make the decisions, with Congress spelling out the broad guiding principles.

III

We are now ready to bring to a head our analysis by concluding that there is an intermediate program of action that offers a way out of our dilemma. It will not give us an immediate surcease from our sorrows, but it will reverse the present trend toward piling up larger surpluses at more and more burden on the Treasury and do this in a way that saves our farming people from the disaster and increasing hardships of the extreme Shuman and less extreme Farm Bureau proposals. It will at the same time set in operation a positive program for bringing about the adjustments which are the real solution of our present unbalanced situation. Given such a program, definite improvement will begin to show by 1960, and by 1965 agricultural recovery will be well on its way.

But it cannot be too strongly stated that something very different from and very much more expanded than our present Farm and Home Development program will be needed. Altogether the agencies participating in working out the loan or adjustment plans and programs for the individual farms and helping farm families find other

3. These are my calculations, not those of the Minnesota authors.

work, — the Co-operative Federal-State Extension Services, the Soil Conservation districts, the Farm Credit Administration, the Farmers' Home Administration, and the Federal Employment agency field staffs — will presently need a total of three to five workers in the average county. Moreover, these field staffs will need to be rigorously trained for and drilled in the performing of their tasks. A beginning workshop for them is just a starter. They will need to be called back after a month for a review of their experience and again after a few more months, and at intervals thereafter. Much of the personnel for this will have to be recruited from college trained men now on farms or in related off-farm work.

Equally essential is it that the agricultural college and experiment station staffs back up this adjustment planning and execution by assembling the data needed for budgeting analysis of alternatives, particularly of alternatives that spread the farm operations over more land and raise net incomes by lowering farm expenditures per acre. This research should specifically include such analysis of alternatives on representative farms — farms representative of different sizes, such as family size, two-man, etc.; also of different combinations of crops and livestock, of different combinations of cultivatable and rolling land, of different soil types, and above all of different leasing arrangements. Such analysis will furnish guidance to the field staffs in the counties that are doing the actual loan or adjustment planning. It scarcely needs to be pointed out that these analyses need to be kept up-to-date to fit changing technology, changing market situations, and the like.

Handling this assignment over the next few years will take a large part of the available research time of a large fraction of the agricultural staffs of the land-grant universities. The few production economists on these staffs will find themselves spreading their efforts as co-workers or consultants over a wide range of projects.

Is the emergency in agriculture so serious that the staffs of our land grant colleges and all the other agencies named should interrupt their regular activities in this way? My answer is, "Yes." Otherwise, we would not be having such extreme measures proposed to us as those I have outlined — making agriculture into a public utility, imposing rigorous quantity quotas on each farm, organizing farmers into unions, and the like.

It will now be apparent that what is here proposed is in effect a mobilization, and strengthening, of the effort of most of the federal, state, and local agencies now set up to serve agriculture, and through agriculture the nation, for the purpose of solving the problem of

mounting agricultural surpluses and hardship for a large fraction of our farm families.

Surely the reader at this point will want to know how the lines of action just designated fit in with those in the President's message to Congress on January 16, and in his veto message of March 31. There are no large differences between them. Both propose a lowering of price-support levels. The President's message would simply give Secretary Benson freedom to lower them as he sees fit down to 60 per cent of the modernized parity level, with no additional assistance to farmers in adjusting to the lower levels set and no compensatory payments. What is here proposed is, for most farm products now under price supports, essentially a gradual lowering of the price supports to a level where supply and demand will come into balance, but along with this a set of gradually lowering supplementary income payments to assist farmers in making the adjustments to the lower price levels, plus a vigorously stepped-up farm planning and loan program as further assistance. Moreover, the setting of prices is not to be left to the Secretary of Agriculture, but this, along with setting the level of supplementary payments, is to be directed by Congress and implemented by a special noninterest board. It will be at once apparent to the reader that these differences could be very important should the current recession extend itself much further.

The other major difference relates to production levels. The veto message stresses not only the unfreezing of the acreage quotas so that producers will be more free to expand their output, but even proposes an actual increase in planted acreages. This would seem to assume that the reduction in stocks in government hands of the past year will be continued. This is to be achieved mainly by increasing further the exports and other disposal of farm products abroad. The procedure of letting the trade handle this disposal, thus keeping the product out of government lands, that was extensively used in the past year, is to be employed still more and applied to other products. It is the judgment of the author that except as our government stands ready to subsidize the underselling of still other exporting nations, any expansion of disposal abroad will have to come about mostly by seeking out and making liberal deals with more nations with large underfed groups. Are we prepared to keep on feeding these underfed groups indefinitely? If not, then the feeding we do now must prepare them to feed themselves better in the near future.

It follows, therefore, that any reasonable program of surplus disposal abroad, plus the other measures of very minor effect outlined

in the veto message — a slight expansion of the school milk program, “market development in co-operation with producer organizations and the food trade,” and “research to find new uses for farm products” — will not keep surpluses sufficiently in control without declines in prices too painful and burdensome for a million or so of our farm families to endure; and that therefore now is not yet the time for any general relaxing of producer quotas.

This last conclusion, of course, assumes no severe droughts or a war in the years just ahead; or even any sustained developments such as those that have produced the temporary rise in farm prices beginning in February and likely to be reversed as early as August.

THE RATE OF INTEREST IN A DYNAMIC MODEL

By A. LL. WRIGHT*

I. Introduction, 327. — II. The Hicksian analysis, 328. — III. The integrated model, 332. — IV. The properties of the integrated model, 338. — V. Conclusion, 343. — Mathematical Appendix, 344.

I. INTRODUCTION

Before 1936 it was generally accepted that prices and the rate of interest were of cardinal importance in determining the level of employment and of income. With the advent of *The General Theory*, however, the emphasis was thrown on to the marginal propensity to consume and the multiplier. Keynes himself, of course, stresses the role of money and interest (although only in a stationary model), and for a while the rate of interest assumed a dominant position. Nevertheless, to the innovators, building models entirely in real terms, even the rate of interest ceded its role to the accelerator. Supply and demand theory, as analyzed in the textbooks, was completely ignored when the business cycle was studied: everything was presented in real terms. A partial bridge between the more conservative elements and these radicals was constructed, in the United Kingdom, in 1950 by Professor Hicks and, in effect, we were all invited to cross over the bridge.

Hicks presented the essence of his theory in real terms, ignoring prices and the rate of interest and substituting in their place the multiplier and accelerator. Although prices never appear in his analysis at all, Hicks does introduce, *ad hoc* so to speak, an essentially stationary theory of the rate of interest which he grafts on to his dynamic model to explain the place of the monetary factor. Now this method of attack has some drawbacks. The interest theory does not form an *integral* part of the model and it does not follow that the real and monetary analysis are mutually consistent with one another. The first part of this article tries to analyze what I think is an important conflict in the two approaches. The second part then aims at constructing a model in which the rate of interest and money form an essential part. Finally, the results which follow from the various assumptions regarding the nature of the money impact on the system are then analyzed.

* St. Salvator's College, St. Andrews.

II. THE HICKSIAN ANALYSIS

The main properties of the basic Hicksian model¹ are now well known, but one or two points connected with it need emphasis. Firstly, there is the familiar consumption function equation, where current real consumption expenditure,² C_t , is made to depend on the real income of the previous period, Y_{t-1} , by means of the marginal propensity to consume, c_Y . Secondly, current real investment expenditure, K_t , is regarded as a linear function of the previous change in real income, $(Y_{t-1} - Y_{t-2})$: the relationship between the two sets of variables is, of course, given by the accelerator, k_Y . Current real income is obtained by adding current real consumption and current real investment.

It will be helpful at this point to introduce a flow diagram of Hicks's model. The diagram, shown in Figure I, is based on that developed by Professor Allen.³ The current variables and the datum are placed in circles and the lines joining the circles show, with the aid of arrows, the direction of dependence between the variables. A small open circle shows where two flows meet, the magnitude of the resulting flow being shown by "+" and "-" signs. Two types of boxes are placed on these lines. Firstly, there are the boxes showing the operators E^{-1} or E^{-2} : this signifies that the dependent variable is connected by a lag of one or two periods to the independent. Secondly, there are the boxes containing the coefficients of the system.

The variables, as in the Hicksian analysis, are all measured in real terms. Current income, Y , is shown in the circle on the top line of the diagram. Proceeding to the left we see that current consumption, shown by C in the circle, depends on the income of the previous period, shown by E^{-1} in the box, by means of the marginal propensity to consume, c_Y , also shown in a box. Going downwards from the Y circle, we see that current investment, K , depends on $(Y_{t-1} - Y_{t-2})$ by means of the accelerator, k_Y . Current consumption and current investment flow on to meet in the open circle on the bottom line of the diagram, where they are added together and flow on to give cur-

1. See J. R. Hicks, *A Contribution to the Theory of the Trade Cycle*, pp. 170-86. Hicks's analysis is, of course, more detailed than the sketch given here, where the aim is only to present the bare bones of his model for purposes of comparison with the model to be constructed in the next section.

2. All the symbols, of course, measure deviations from an assumed equilibrium level (which may be zero), but this is immaterial for the discussion presented here.

3. See R. G. D. Allen, "The Engineer's Approach to Economic Models," *Economica*, NS: XXII (1955), 158-68, and R. G. D. Allen, *Mathematical Economics* (London, 1956), pp. 284-89. The notation used above is slightly different from that adopted by Allen.

rent income, thus completing the circuit of the system. The exogenous injection, X , is shown in the circle at the bottom of the diagram. This circle is not connected to the rest of the flow system, showing that the impact is merely a single nonrepeated shock. If X was continuously repeated in every period a continuous line would join the X circle to the rest of the system.

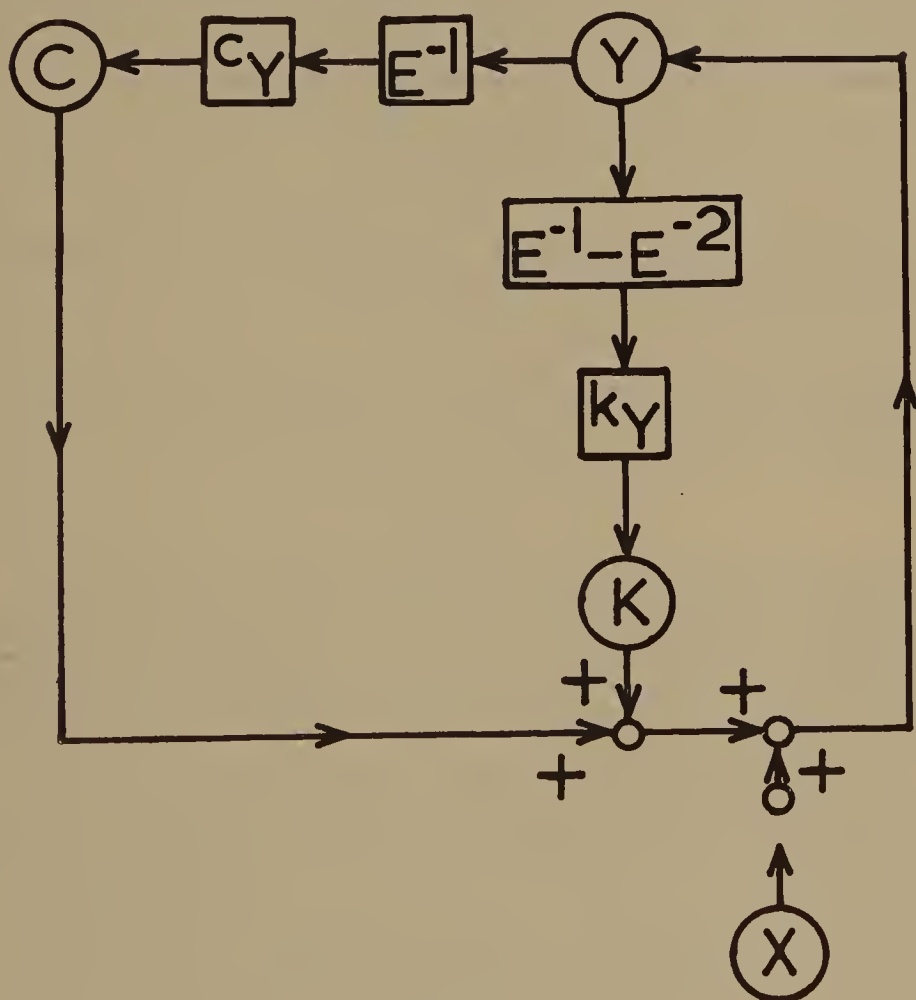


FIGURE I
THE BASIC HICKSIAN MODEL

The dynamic properties of this system are quite familiar. Depending on the values of the marginal propensity to consume and of the accelerator, income can diverge (either directly or in cycles of ever-increasing amplitude) or converge (again, either directly or in cycles of ever-decreasing amplitude). Even more obvious but, unfortunately, far more neglected are what might be termed the stationary

properties of the system. Here the characteristics can be most easily framed in *ex ante* and *ex post* terms, using these concepts in the now familiar way and most clearly employed by Mr. Alexander in his recent analysis of the Harrodian growth model.⁴

Notice, as a preliminary point, that both consumption and investment plans are completely fulfilled: *ex ante* consumption equals *ex post* consumption and *ex ante* investment equals *ex post* investment. If, then, the community has a received income of Y_{t-1} and plans to consume C_t in the next period, it must plan to save $Y_{t-1} - C_t$. In other words, the *ex ante* saving of period t , S_t^a , is given by

$$S_t^a = Y_{t-1} - C_t.$$

The *ex post* saving of period t , S_t^p , on the other hand, must equal, by definition, the investment of period t , so that

$$S_t^p = Y_t - C_t.$$

Obviously, then, *ex ante* saving only equals *ex post* saving (i.e., investment) if

$$Y_{t-1} - C_t = Y_t - C_t$$

or

$$Y_{t-1} = Y_t.$$

In other words, income must be constant for the equality of *ex ante* saving and investment, while income will be increasing if investment is greater than *ex ante* saving and decreasing if the latter is greater than the former. This, of course, merely conforms to the accepted usage of these terms.

Now Hicks's theory of the rate of interest⁵ relates "saving," investment, the rate of interest and income together, but he is nowhere explicit as to whether it is *ex ante* or *ex post* saving which should figure in this analysis. In his discussion in his "Trade Cycle" it is difficult to decide which has been given the role. "If like Keynes," we are told, "we neglect any direct influence of interest on saving, we can go on to say that with a given value of investment . . . the equilibrium value of money income will be determined by the multiplier."⁶ Since it is the rate of interest that helps determine investment, it is thus possible to relate the level of income to the rate of interest. Now it is vitally important to notice the steps in this argument: the multiplier shows the relationship between "saving" and

4. See S. S. Alexander, "Mr. Harrod's Dynamic Model," *Economic Journal*, LX (1950), 724-39.

5. See J. R. Hicks, *op. cit.*, chap. XI, pp. 136-54.

6. *Ibid.*, p. 137.

income — this follows from the definition of the multiplier; and *it is because this "saving" is made equal to investment* that investment and income can be related by the multiplier; so that, finally, income becomes a function of the rate of interest.⁷ This, of course, is just what Hicks has done in constructing what he calls the *SI*-curve — a downward sloping curve showing the relationship between income and the rate of interest. It might thus appear that it is *ex post* saving rather than its *ex ante* counterpart that figures in the analysis. However, Hicks now relaxes some of his assumptions. "If a fall in the rate of interest makes the representative individual save less out of a given income . . . the *SI*-curve is made more elastic; if a fall in the rate of interest makes the representative individual save more, the *SI*-curve will become less elastic."⁸ Now since Hicks agrees that investment (i.e., *ex post* saving) and the rate of interest are negatively correlated he must here be talking about *ex ante* saving.⁹ He now transfers this interest-income framework into a dynamic system where he allows lags to affect the simple operation of the model. This, however, is in no way permitted to influence the *structure* of the *SI*-curve: this curve, built on the basis of a stationary system (i.e., in short, a system in which *ex ante* saving equals investment), is utilized in the dynamic model without any alteration.

It is just at this point that both Hicks's analysis and the dynamic model that I attempted to construct¹ get into difficulties. Turning back to the Hicksian dynamic model, we saw there that *ex ante* saving only equals investment when income is constant. Consequently, this analysis of saving and the rate of interest based on the *SI*-curve is only valid if income does not change. The whole core of the analysis is invalidated by a nonconstant income since the *SI*-curve, based as it is on the fact that *ex ante* saving equals investment, can no longer be constructed.

7. The steps are more explicitly set out by Hicks in his earlier presentation of the rate of interest theory: see J. R. Hicks, "Mr. Keynes and the 'Classics': a Suggested Interpretation," *Econometrica*, V (1937), 149. Using the symbols to represent absolute values rather than deviations, we have $Y = S + C$, or $dY = dS + dC$; but since $C = c(Y)$, so that $dC = c_Y dY$, it follows immediately that $dY = dS/(1 - c_Y)$. Now $K = k(r)$ so that $dK = k_r dr$. Assuming that $dS = dK$ we get immediately that $dY/dr = k_r/(1 - c_Y) < 0$, which gives the slope of Hicks's *SI*-curve. It is this assumption that $dS = dK$ in a dynamic analysis that is suspect: see A. Ll. Wright, "Sequence Analysis and the Theory of the Rate of Interest," *Economic Journal*, LXV (Dec. 1955), 630-31.

8. J. R. Hicks, "Contribution to the Theory of the Trade Cycle," p. 139.

9. Again referring to Hicks's earlier article (*Econometrica*, V (1937), 149) it will be more clearly seen that it is *ex ante* saving which enters the analysis. At any rate, if *ex post* concepts were used in a supply and demand analysis the whole force of the supply effects would be nullified in a rate of interest analysis.

1. See A. Ll. Wright, *loc. cit.*, pp. 626-40.

In my previous analysis of this problem I followed Hicks and let "saving" equal investment. From a purely mathematical point of view² the system is overdeterminate: Hicks did not attack this problem although I attempted to solve it by introducing an additional (short-term) rate of interest. From an economic point of view such "saving" must include both *ex ante* saving and changes in stocks. If income is decreasing the decrease over any two periods can be regarded as an increase in stocks.³ Subtracting this increase from *ex ante* saving we must get the "saving" which is equal to investment. Similarly, if income is increasing the decrease in stocks can be added on to *ex ante* saving to give the "saving" that equals investment.

Both ways to this escape must be admitted to be artificial. The problem of integrating the monetary and real aspects of the system has a solution, but a reasonable solution can be obtained only if the two following points are clearly recognized: firstly, that it is *ex ante* saving that has the role to play in the rate of interest analysis and, secondly, that such *ex ante* saving need not necessarily equal *ex post* saving (or investment). On this basis an integrated model can now be constructed.

III. THE INTEGRATED MODEL

Probably the easiest way of explaining⁴ how all the variables of the integrated model are connected together is by means of the flow diagram shown in Figure II. As prices have no part to play, all the variables are measured in real terms. This, of course, is a serious limitation and a strong point of criticism against the approach, but since it is not the purpose here to attack this kind of simplification it can be ignored. Secondly, it must be noticed that in strict logic the variables of the model are measured as deviations from their values in the original (equilibrium) position of the initial period. However, this assumption presents no limitation for the purposes in hand: at any rate, the initial period can always be so selected that the values of the variables were then zero.

First of all, there is consumption expenditure. This is determined

2. It need hardly be pointed out, of course, that this problem of overdeterminateness is not just a mathematical illusion. From an economic point of view the failure to solve it involves two different solutions for the value of investment.

3. This is a rather heroic definition of "stocks," despite the fact that it seems to be one which figures from time to time in some of the Keynesian type of analyses. Since both *ex ante* equals *ex post* consumption and *ex ante* equals *ex post* investment, it is hard to discover what kind of stocks are meant to be changing in volume.

4. The basic equations of the model are presented in section 3 of the Mathematical Appendix: see *infra*, pp. 346-48.

in the same way as in the Hicksian model: current consumption depends on lagged income by means of the marginal propensity to consume and is shown by the flow rising out of the Y circle to the left and meeting the C circle. Again, just as in the Hicksian model, current *ex ante*⁵ saving, S , is obtained by subtracting current consumption from lagged income. In the diagram current consumption is subtracted from lagged income in the small open circle immediately

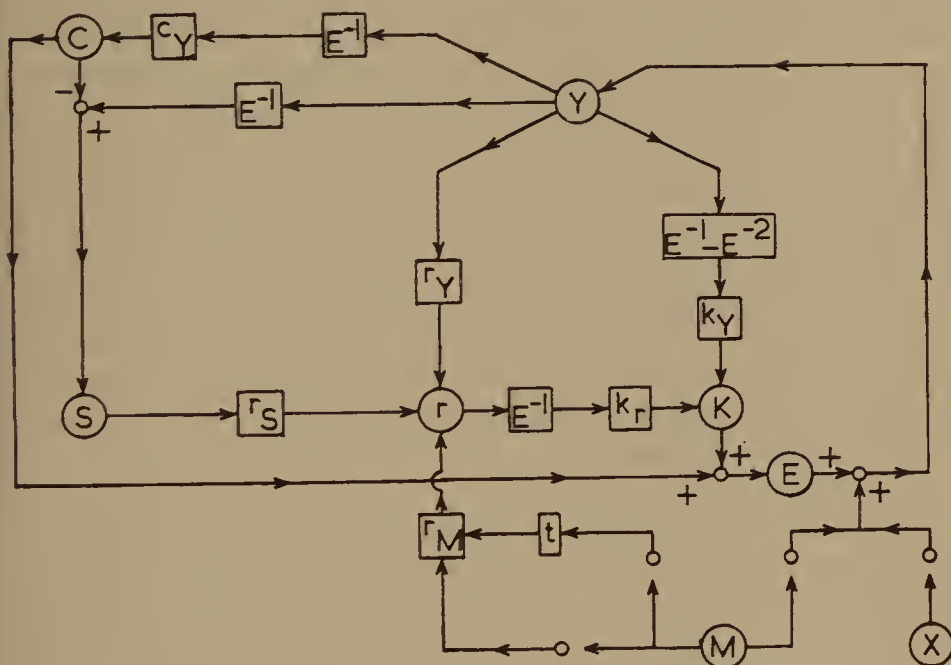


FIGURE II
THE INTEGRATED MODEL

below the C circle and flows down to give current *ex ante* saving in the S circle.

The rate of interest, r , has no part to play in the Hicksian model and must, therefore, be examined in some detail. In the first place, following Hicks's analysis in "Value and Capital,"⁶ where prices have an important part to play, it can be assumed that the rate of interest is constant during any period, altering only at the *end* of each period (in the Hicksian terminology of "weeks," or "Saturdays"). During the *current* period, then, it is the rate of interest of the *previous* period, determined by forces operating in this previous and earlier periods,

5. Since we are mainly concerned with *ex ante* saving, the superscript "a" can be dropped.

6. See J. R. Hicks, "Value and Capital," 2d ed., pp. 122-24.

that rules. The *current* rate of interest is only determined at the *end* of the *current* period and remains fixed for the *following* period. Consequently, the current investment market is influenced by the rate of interest of the previous period. Now it is possible to regard the current rate of interest as being determined at the *beginning* of the *current* period (i.e., on "Mondays" — nothing happens on "Sundays"). Then it would be the forces of the previous period(s) that determine its value while it, in turn, would affect the investment market in the *current* period. This, of course, leads to exactly the same relationships between the investment market and the rest of the economy as in the case of the other formulation. In other words, only a purely terminological change is involved in the selection of the "day" on which the rate of interest changes. The factors, then, which determine the rate of interest of the previous period are the reactions taking place in the previous period(s), while the previous rate of interest itself influences the reactions which occur during the current period. In the absence of expectations, which are not explicitly brought into this analysis anyway, this is the only sequence of events, apart from additional lagging, that can occur: no changes in the current period can influence the rate of interest set up at the end of the previous period.

For purely pedagogic purposes the forces that determine the rate of interest set up at the end of the current period, rather than at the end of the previous period, can be examined. The first factor influencing it is *current ex ante* saving. Now it is the *willingness* of people to supply savings that influences the rate of interest in such a way that an increased willingness to save reduces the rate of interest. In other words, the larger *current ex ante* saving the lower is the current rate of interest. Hence r_s , the factor showing how, *ceteris paribus*, the rate of interest changes in response to a change in *ex ante* saving, must be negative. This is shown in the flow diagram by the line connecting the r and S circles and containing the r_s box. If *current ex post* saving (or investment) is allowed to affect the current rate of interest instead, the essence of the supply effect on the rate of interest is lost. At any rate, it would be impossible to say which way an increase in *current ex post* savings would affect, if at all, the current rate of interest. Even apart from this, however, in a dynamic analysis it is through income changes, and not through changes in the rate of interest, that the equality of *ex post* saving and investment is attained.

The effect, then, of (say) an increase in the marginal propensity to consume is, *ceteris paribus*, to raise the rate of interest since con-

sumption is increased and *ex ante* saving is reduced. It could be argued, of course, that it is not *ex ante* saving as defined here that affects the rate of interest, and that it is only that part of nonconsumed income that is directly put into the investment market (for old securities or for new issues) which exerts such an influence. In other words, some part of nonconsumed income may be held as money (i.e., hoarding in some sense) and this will not affect the rate of interest at all. This criticism, I would agree, is valid, but I would maintain that the effect set up by such an adjustment is negligible in the ordinary course of events, and can, therefore, be safely neglected. In addition, it would be exceedingly difficult to determine the proportions into which *ex ante* saving would have to be divided.⁷ Rather than complicate the model by bringing in a term to allow for such small changes in hoarding as might occur, it is easier to suppose that the value of the r_s coefficient is altered in such a way that a tendency to increased hoarding is reflected in a numerically smaller value of it.

Consequently, the model allows *ex ante* saving, or the willingness of people to save, to affect the rate of interest. On the other hand, the rate of interest itself has not been allowed to affect the willingness of people to save. It would be possible, of course, to connect current consumption to the rate of interest of (say) the previous period, thus allowing for a rate of interest effect on consumption. Technically, this would mean that the final solution of the system was given by a cubic equation rather than by the simpler quadratic form. Now, firstly, taking the community as a whole, it is not certain whether a change in the rate of interest would alter consumption in the same or in the opposite direction. Secondly, whatever influence the rate of interest does, in fact, have on consumption it would be exceedingly small in comparison with the other effects in the system. On balance, therefore, nothing of fundamental importance is lost if the rate of interest is not allowed to affect the willingness to save while the latter is allowed to affect the former.

The second factor influencing the current rate of interest is *current* income, and is shown by the flow connecting the r and Y circles. The current rate of interest, it will be recalled, is set up at the *end* of the current period, and the simplest income effect to introduce is that of income changes in the current period. This element enters along purely conventional Keynesian lines, bringing in the effect of

7. It is here, of course, that the purely speculative motive for holding money enters the system, but it is not my intention to introduce it into the simple model presented in this article.

the transactions motive to liquidity. r_Y , the coefficient giving the relationship between the two variables, is, therefore, positive: an increase in income, *ceteris paribus*, raises the rate of interest.

The third factor affecting the rate of interest is money, M (including both cash and deposits). Money can be regarded as a datum of the model given by the banking system or government: for this reason it is placed by itself in a circle at the bottom of the diagram. Now the rate of interest, as in Keynesian theory, depends on the *total* amount of money in the system. The larger is the amount of money the lower is the rate of interest: hence r_M , the coefficient showing the relationship between changes in the amount of money and changes in the rate of interest, must be negative. The model, it will be recalled, is framed in terms of deviations of the variables from their values in period 0 and this means that the current rate of interest must depend on the *total* volume of money that has been created since period 0. Suppose that in period 1 the economy has been subjected to a single injection of money and that this injection has not been repeated in any subsequent period. The system has experienced what may be termed a *single nonrepeated inflationary injection*.⁸ Money rises by this amount, say M , and does not deviate any further than by M units from its initial (equilibrium) level. Hence, the rate of interest must depend on M . The M circle is connected to the r circle by means of the lower circuit emerging from M . Secondly, however, the injection of money, of (say) M per period, may be continuously repeated in each period. This can be called the *continuously repeated inflationary injection* case. Here the deviation of money from its equilibrium level is being continuously enlarged and the current rate of interest thus depends on an increasing quantity of money. The current period can be called period t , and since the quantity of money is increasing by M units in each period the deviation of money in period t from its value in period 0 must be tM . Consequently, in this case, the current rate of interest depends, not just on M , but on tM . The M circle is connected to the r circle by means of the upper circuit containing the box t . Finally, of course, no new money may be created at all so that the M circle and its circuit can be completely ignored.

Investment demand follows the Hicksian pattern. The main determinant of current investment is felt through the change in income over the two previous periods by means of the accelerator, k_Y .

8. It will be noticed that "inflationary" is used here merely as a convenient adjective and does not refer to an inflationary situation in the proper sense of that term.

The circuit coming down from the Y circle and linking it to the K circle shows this effect. The distinguishing feature here is the influence which the rate of interest of the previous period is allowed to exert on current investment, shown by the connection between the r and K circles. Obviously a higher previous rate of interest must, *ceteribus paribus*, reduce current investment, so that k_r , the coefficient showing this relationship, is negative.

Total current expenditure, E , is obtained by adding current consumption and current investment. A flow from the C circle is added to the flow from the K circle and joined to the E circle.

The impact of the exogenous injection enters the system between expenditure and income.⁹ Here four cases, shown at the bottom of the flow diagram, can be distinguished. Firstly, there are the two *noninflationary* cases, as they may be called: the impact to which the system is subjected is not financed by the creation of any new money at all, so that the exogenous injection has no direct effect on the rate of interest. First of all, this injection may take the form of a *single nonrepeated noninflationary injection*, X . The flow from the X circle is felt just once, and joins the expenditure-income circuit for the initial period only, providing merely a starter to the system. On the other hand, there may be a *continuously repeated noninflationary injection* of X units in every period.¹ Here the flow from X is permanently linked to the rest of the flow system. The other two cases are those where newly created money is pumped into the system — the *inflationary* cases which were discussed in connection with the rate of interest. With the *single nonrepeated inflationary injection* of M , the M circle, in addition to influencing the rate of interest along the lower branch to the left of the M circle, is also connected once only to the system in the expenditure-income circuit. Finally, with a *continuously repeated inflationary injection* of M units in each period a permanent flow is set up from M to the expenditure-income circuit while, as has been seen, an ever-increasing effect is exerted by money on the rate of interest through the left-hand upper branch from the M circle.

A comparison of Figures I and II shows just how strongly the

9. Since price effects are being deliberately neglected, it makes no difference, of course, whether the exogenous injection enters the system before or after expenditure.

1. The model does not take account of government revenue (a leak) or government expenditure (an injection) so that in this case the injection must be imagined to be financed by budget surpluses accumulated before period 0. The case is, therefore, more in the nature of a mere theoretical possibility than of any practical significance.

integrated model relies on that of Hicks. The same characteristic connections between income, consumption and investment are evident. The differences, too, are clearly brought to light. In the second model it is the series of flows which determine the rate of interest and, in turn, influence investment that are important. The element, as we will see in the next section, that makes for a permanent expansion of income, irrespective of the relative values of the coefficients and the size of the flows between income, consumption and investment, can be clearly seen: a continuously repeated injection of money must have an increasing influence on the system as t grows. That there is no such influence in the Hicksian system is patently shown by the flow diagram of Figure I.

IV. THE PROPERTIES OF THE INTEGRATED MODEL

The Stationary Properties. When the exogenous injection takes the form of a single nonrepeated impact, current expenditure equals current income and it can be easily seen, just as in the case of the basic Hicksian model, that income is increasing, constant or decreasing according as *ex ante* saving is less than, equal to or greater than *ex post* saving (or investment). On the other hand, when the exogenous injection is continuously repeated in each period income becomes greater than expenditure. Following the usual practice,² however, the exogenous injection can be regarded as part of investment, and the same conclusions hold with respect to the relationship between *ex ante* and *ex post* saving.

The Permanent Effects of the Exogenous Injections. In the integrated model, just as in any model based on difference equations, the elements that contribute to the change in income can be split into two parts. Firstly, there is the permanent effect which the exogenous injection itself exerts. Secondly, there is the effect which the *relative* values of the coefficients exert, giving certain cyclical, explosive or asymptotic characteristics to the income time paths. To explain the characteristics of the system these two effects can be analyzed separately and then combined to give the final picture.

The first task is to determine the permanent effect of the exogenous injection on income. In the case of the single nonrepeated non-inflationary injection the system experiences no permanent effect from the exogenous injection. The sole function of the latter is to provide a "starter" to the system. In terms of Figure III, this means that

2. See, e.g., Joan Robinson, "Introduction to the Theory of Employment," pp. 16-18.

the exogenous injection leaves income at zero on the t axis (i.e., at the original level).

In the other three cases the broad characteristics of the permanent effects of the exogenous injections are determined independently of the values of the coefficients, although the actual magnitudes of the permanent effects do depend, of course, on these values. Assuming a convenient set of values for the coefficients³ in order to simplify the verbal presentation, the magnitude of the effects can be shown quite simply in Figure III.



FIGURE III

THE PERMANENT EFFECT OF THE EXOGENOUS INJECTION

O — the original equilibrium level of income

First of all, there is the case of the continuously repeated non-inflationary injection. Here the magnitude of the accelerator has no effect on the permanent influence of the exogenous injection at all: the main determining factor is the marginal propensity to consume. With the latter having a value of 0.9, and the other coefficients having the given values, the permanent effect of an exogenous injection of 1 unit is to raise income by just under 10 units:⁴ this is shown by the curve marked 2 in Figure III. In the case of an ordinary multiplier

3. The equations giving the particular solutions of the general difference equations are to be found in the Mathematical Appendix. It is assumed in the text that the relevant coefficients have the following values: $k_Y = 2.0$; $k_r = -0.1$; $r_S = -0.1$; $r_Y = 0.05$ and $r_M = -0.1$.

4. The equation is given by

$$\bar{Y} = z_1 = \frac{2X}{1.99 - 1.98c_Y}.$$

series, a marginal propensity to consume of 0.9 would raise income to exactly 10 units. If the marginal propensity to consume falls to 0.8, and the other coefficients remain the same, income would be raised to slightly under 5 units: again, a result almost identical to that found for a simpler multiplier series is obtained. In other words, the permanent effect of the exogenous impact in the case of a single nonrepeated noninflationary injection is much the same as that obtained from the simpler kind of multiplier model, being magnified by a larger marginal propensity to consume. With a single nonrepeated inflationary injection the value of the accelerator has, again, no effect. The permanent effect of an impact of 1 unit of newly created money is to raise income to one-hundredth of the level obtained in the previous case. Consequently, if the marginal propensity to consume is 0.9 then income is raised by 0.1 units to the line marked 3 in Figure III: if the marginal propensity to consume falls to 0.8 income rises by 0.05 units. If the values of the coefficients are varied (within reasonable limits, of course) the characteristics of these results are not altered. With a continuously repeated noninflationary injection income is always *raised*, and that by a multiple of the exogenous injection: with a single nonrepeated inflationary injection income is also always *raised*, but only by a fraction of the injection. It is this kind of characteristics, which are virtually independent of the values of the coefficients, that are important.

The most interesting case is that of the continuously repeated inflationary injection. In the three previous cases the permanent effect of the exogenous injection resulted in a constant level of income: here there is the significant difference that income *increases continually* at a constant rate. The marginal propensity to consume is again the main determinant of the rate of increase, the accelerator having no part to play⁵ although its value does affect the *level* at which income grows. Given the values of the other coefficients, a marginal propensity to consume of 0.8 results in a 5 per cent rate of growth per period: an increase to 0.9 raises the rate of growth to 10 per cent. This is shown as line 4 in Figure III: with an impact of 1 unit of newly created money per period income starts its increase slightly above the level reached in the case of the continuously repeated non-

5. The equation is given by

$$\bar{Y} = z_3 + z_4 t = M \left[\frac{3.94 + 0.04k_Y - 3.96c_Y}{(1.99 - 1.98c_Y)^2} + \frac{0.02}{1.99 - 1.98c_Y} t \right]$$

so that the rate of increase of income is given by

$$\frac{d\bar{Y}}{dt} = z_4 = \frac{0.02}{1.99 - 1.98c_Y}$$

inflationary injection. The important point to notice here is that, no matter what the values of the coefficients are, the effect of the exogenous injection is always to increase income continuously at a constant rate. This, of course, is a sensible result and it is obtained because the newly created money has been allowed to affect the rate of interest.

The Cyclical and Other Characteristics. The next problem is to examine, rather more briefly, the effect of the relative values of the

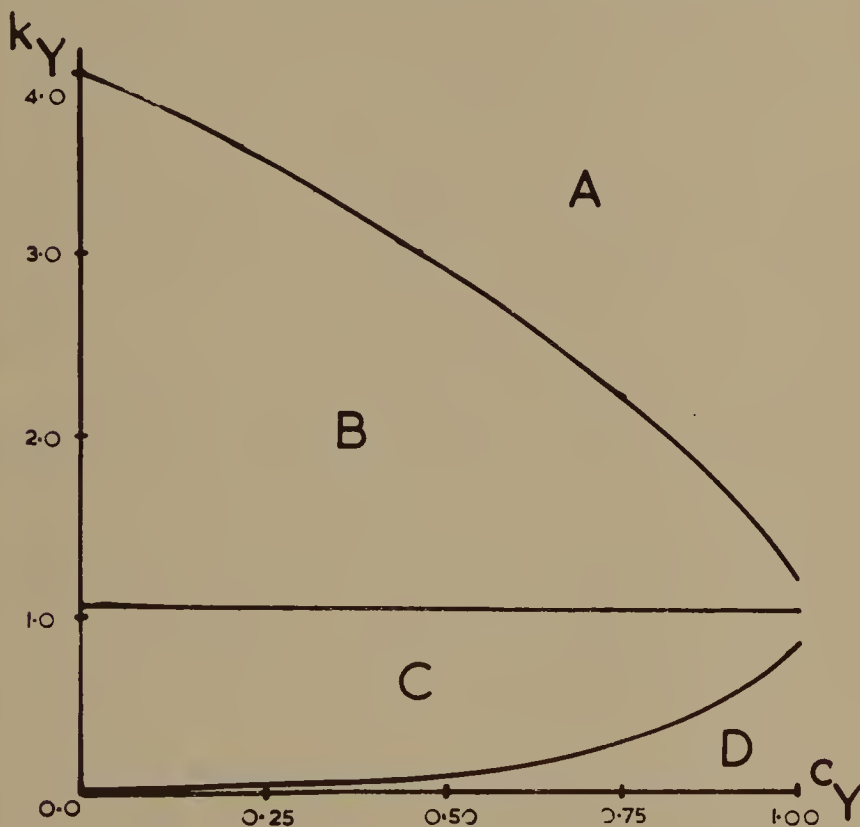


FIGURE IV

THE CYCLICAL CHARACTERISTICS

coefficients. Here the results are the same for all four types of exogenous injections. The relative values of all the coefficients, except the one showing the relationship between money and the rate of interest, enter into this analysis. The most important, however, are the accelerator and the marginal propensity to consume, and the main properties of the system can be most easily examined by letting the values of these two coefficients alone vary and by assuming that the values

of the others are constant.⁶ Again, a graph,⁷ shown in Figure IV, helps considerably in presenting the analysis.

The graph shows how the main characteristics of the system which are relevant here depend on the values of the marginal propensity to consume and the accelerator. Each point in the diagram must lie in one of four regions and represents a selection of values of the marginal propensity to consume and the accelerator, while the characteristic development of income over time is determined by the region into which the point falls. Firstly, there is region A with, broadly speaking, relatively large values of the accelerator. In this case income will expand relentlessly from its initial (equilibrium) value. Secondly, there is region B with slightly smaller relative values of the accelerator: here income will develop in cycles of ever-increasing amplitude. Both these cases are explosive, and for their existence the accelerator must, in broad terms,⁸ be greater than unity. In the other two cases⁹ the accelerator is less than unity. With a point lying in region C cycles of ever-decreasing amplitude are set up. Finally, with relatively small values of the accelerator the resulting income time path depends on the nature of the exogenous injection.

The Combined Results. Combining the two groups of effects we find that the simplest case is that where the relation between the marginal propensity to consume and the accelerator lies in either region A or B. No matter what form the injection takes income continuously diverges from its original (equilibrium) level. For region A the divergence is noncyclical: in the two noninflationary cases and that of a single nonrepeated inflationary injection income expands at the same constant *rate*, although at the lowest *level* in the case of the single nonrepeated noninflationary injection and at the highest *level* when the noninflationary injection is continuously repeated. With a continuously repeated inflationary injection income expands at a continuously increasing rate. With the explosive cyclical cases of region B the cycles are generated around the four curves of income shown in Figure III. In the other two cases of regions C and D, income even-

6. It is assumed that they have the values given in footnote 3, p. 339.

7. This graph follows that of Professor Samuelson: see P. A. Samuelson, "Interactions between the Multiplier Analysis and the Principle of Acceleration," *Review of Economic Statistics*, XXI (May 1939), 75-78, reprinted in *Readings in Business Cycle Theory*, ed. G. Haberler, pp. 266-69.

8. More exactly, $k_Y > 1 + (1 - c_Y)/100$: this is in contrast to the Hicksian model where $k_Y > 1$ is a sufficient condition for this type of explosion.

9. In the limit, of course, the point may lie on the line separating regions B and C: here cycles of constant amplitude are set up. This case is, however, merely a theoretical curiosity since there are no a priori reasons why the coefficients should become stuck at these values.

tually returns to the line set up by the permanent effect of the exogenous injection. This means that with a single nonrepeated noninflationary injection income returns to its original (equilibrium) level. In the cases of the continuously repeated noninflationary and single nonrepeated inflationary injections income expands to the limits set by the lines in Figure III. For region C the damped cycles are set up around these lines. The significant case is that of the continuously repeated inflationary injection: here income eventually settles down to a steady rate of growth, shown by line A in Figure III.

The properties of the model can be most easily summarized in a table. The most important points that emerge are that in all cases

THE INCOME TIME PATHS

	Single nonrepeated noninflationary injection	Continuously repeated noninflationary injection	Single nonrepeated inflationary injection	Continuously repeated inflationary injection
Region A	Constant rate of increase	Constant rate of increase	Constant rate of increase	Increasing rate of increase
Region B	Explosive cycles around a zero level	Explosive cycles around a nonzero constant level	Explosive cycles around a nonzero constant level	Explosive cycles around a rising level
Region C	Damped cycles around a zero level	Damped cycles around a nonzero constant level	Damped cycles around a nonzero constant level	Damped cycles around a rising level
Region D	Zero level	Constant non- zero level	Constant non- zero level	Constant rate of increase

except that of a continuously repeated inflationary injection income will *only* expand if the accelerator is greater than unity: but in this exceptional case, no matter what the values of the coefficients, income moves further and further away from its original (equilibrium) level.

V. CONCLUSION

The rate of interest, then, has quite a significant part to play in the more elaborate model. Indeed, the attempt to analyze investment entirely independently of the rate of interest would seem to neglect completely one of the important channels through which money exerts an influence on the system. The similarity between the conclusions reached here and those based on the Keynesian analysis are,

of course, obvious: nor is this a mere coincidence, since the Keynesian framework for this mechanism has been incorporated into the dynamic model.

The analysis, however, poses another, and possibly more important, category of problems. It has been assumed throughout that the variables have been measured in real terms. Now obviously income cannot go on increasing indefinitely in real terms in response to continuous injections of newly created money. There must be some limit to the height to which real income can rise and this limit is set mainly by the state of technological development. It is possible, of course, to fix a ceiling to real income as Hicks has done. However, once increasing money is allowed to play a part, the assumption of a fixed ceiling to real income development becomes rather dubious. Long before this ceiling has been reached price rises will have taken place. As long as it can be assumed, possibly along the lines of an overdose of money illusion, that price changes do not in any way affect changes in the real values of the variables then the effect of such price changes can be ignored. Even if a pure inflation develops, with a constant real income and an ever-increasing money income, an analysis framed solely in real terms may be of some use. However, if the price changes themselves are permitted to influence the real values of the variables, and this is a realistic assumption, an analysis presented entirely in real terms becomes futile. In such cases it is necessary to have a model in which the real and money changes in the variables interact with one another. It is outside the scope of this article, of course, to pursue these ideas: although the need for such an approach must be emphasized, the construction of such a model must be presented in another story.

MATHEMATICAL APPENDIX

1. *The Symbols*

The variables of the system are measured in real terms as deviations from an original (equilibrium) position in period 0.

Y	income	E	total expenditure
C	consumption	r	the rate of interest
S	<i>ex ante</i> saving	X	noninflationary injection
K	investment	M	inflationary injection

The coefficients of the system are expressed in the notation of partial derivatives: the dependent variable is shown by the small letter and the subscript indicates the independent variable.

c_Y the marginal propensity to consume: $1 > c_Y > 0$

- k_Y the accelerator: $k_Y > 0$
 r_Y the income effect on the rate of interest: $1 > r_Y > 0$
 r_S the *ex ante* saving effect on the rate of interest: $-1 < r_S < 0$
 r_M the money effect on the rate of interest: $-1 < r_M < 0$
 k_r the rate of interest effect on investment: $-1 < k_r < 0$
 $\alpha \equiv c_Y + k_r r_Y$: $k_r r_Y < 0$, but must be numerically far smaller than c_Y ,
 hence $\alpha > 0$.
 $\beta \equiv k_r r_S (1 - c_Y)$: $1 > k_r r_S > 0$, hence $\beta > 0$.
 $1 - \alpha - \beta = (1 - c_Y)(1 - k_r r_S) - k_r r_Y > 0$.

2. The Hicksian Model

The basic Hicksian model contains three equations and three dependent variables, Y_t , C_t and K_t , and is, therefore, correctly determinate.

$$C_t = c_Y Y_{t-1} \quad (2.1)$$

$$K_t = k_Y (Y_{t-1} - Y_{t-2}) \quad (2.2)$$

$$Y_t = C_t + K_t = (c_Y + k_Y) Y_{t-1} - k_Y Y_{t-2} \quad (2.3)$$

Hicks's model of the rate of interest,¹ written as deviations (i.e., differentials) in terms of the symbols used here, is given for a stationary economy and can be presented, in effect,² by four equations.

$$r = r_M tM + r_Y Y \quad (2.4)$$

$$K = k_r r \quad (2.5)$$

$$S = (1 - c_Y) Y \quad (2.6)$$

$$S \equiv K \quad (2.7)$$

Here there are four unknowns, S , K , Y and r , and one datum, M ; the system is determinate. Dynamizing this stationary system and combining it with the equations of his basic model, the following system can be obtained.

$$C_t = c_Y Y_{t-1} \quad (2.8)$$

$$K_t = k_Y (Y_{t-1} - Y_{t-2}) + k_r r_{t-1} \quad (2.9)$$

$$S_{t-1} = Y_{t-2} - C_{t-1} \quad (2.10)$$

$$S_{t-1} \equiv K_{t-1} \quad (2.11)$$

$$r_{t-1} = r_M (t - 1)M + r_Y Y_{t-1} \quad (2.12)$$

1. See J. R. Hicks, *op. cit.*, *Econometrica*, V (April 1937), 153.

2. Hicks writes this as a system of three equations with three unknowns, since he regards $S \equiv K$ and so suppresses the S variable altogether. It can also be noticed that he writes the liquidity preference function as $M = L(Y, r)$ rather than as (2.4), but this is immaterial for the purposes in hand.

$$\text{or} \quad r_{t-1} = r_M(t-1)M + r_Y Y_{t-1} + r_S S_{t-1}$$

$$Y_t = C_t + K_t \quad (2.13)$$

In other words, there are six independent equations, but only five dependent variables; viz., Y_t , C_t , K_t , S_{t-1} and r_{t-1} ; and one datum, M . The system is overdeterminate. The cause of this stems alone from the dynamizing of the system. In the model presented here this is essentially evidenced by the lagging of income, and it is the lagging of this variable, particularly in equations (2.8), (2.10) and (2.13), (and not in the rate of interest equation) that creates the difficulty.

In my previous article³ I introduced an additional (short term) rate of interest bringing the number of dependent variables up to six. Now I think, for the reasons stated in the text, that S_t should be regarded as *ex ante* saving and not as *ex post* saving. Hence the identity (2.11) should be abandoned, so reducing the equations in the above system to five.

3. The Integrated Model

This system follows directly once the adjustment has been made to the set (2.8) to (2.13). Following the order adopted in the text the system can be presented by six⁴ basic equations.

Current consumption:

$$C_t = c_Y Y_{t-1} \quad (3.1)$$

Previous *ex ante* saving:

$$S_{t-1} = Y_{t-2} - C_{t-1} = (1 - c_Y) Y_{t-2} \quad (3.2)$$

Previous rate of interest:

i. Continuously repeated inflationary injection:

$$r_{t-1} = r_S S_{t-1} + r_Y Y_{t-1} + r_M(t-1)M \quad (3.3a)$$

ii. Single nonrepeated inflationary injection:

$$r_{t-1} = r_S S_{t-1} + r_Y Y_{t-1} + r_M M \quad (3.3b)$$

iii. Noninflationary injection:

$$r_{t-1} = r_S S_{t-1} + r_Y Y_{t-1} \quad (3.3c)$$

3. See A. Ll. Wright, *op. cit.*, *Economic Journal*, LXV (Dec. 1955), 628 and 633-34.

4. The extra equation appears because the additional dependent variable for current expenditure, E_t , has been introduced for simplicity of presentation in the later analysis.

Current investment:

$$K_t = k_Y(Y_{t-1} - Y_{t-2}) + k_r r_{t-1} \quad (3.4)$$

Current total expenditure:

$$E_t = C_t + K_t \quad (3.5)$$

Current income:

i. Single nonrepeated noninflationary injection:

$$Y_t = E_t \quad (3.6a)$$

ii. Continuously repeated noninflationary injection:

$$Y_t = E_t + X \quad (3.6b)$$

iii. Single nonrepeated inflationary rejection:

$$Y_t = E_t \quad (3.6c)$$

iv. Continuously repeated inflationary injection:

$$Y_t = E_t + M \quad (3.6d)$$

In this system there are six independent equations, (3.1) to (3.6), selecting one each from the sets (3.3) and (3.6). There are six dependent variables, viz., Y_t , E_t , C_t , K_t , S_{t-1} and r_{t-1} ; and one datum, either M or X . The system is thus correctly determinate.

By simple elimination it will be found that, in the most general case of the continuously repeated inflationary injection,

$$E_t = (k_Y + \alpha)Y_{t-1} - (k_Y - \beta)Y_{t-2} + k_r r_M(t - 1)M. \quad (3.7)$$

Depending on the assumptions regarding the nature of the exogenous injection, this gives rise to four equations for determining current income.

i. Single nonrepeated noninflationary injection:

$$Y_t = (k_Y + \alpha)Y_{t-1} - (k_Y - \beta)Y_{t-2} \quad (3.8)$$

It will be noticed immediately that in the Hicksian model $k_r = 0$, so that $\alpha = c_Y$ and $\beta = 0$; in the case of (3.8), then, these limitations reduce the equation to that of Hicks's basic model.

ii. Continuously repeated noninflationary injection:

$$Y_t = (k_Y + \alpha)Y_{t-1} - (k_Y - \beta)Y_{t-2} + X \quad (3.9)$$

iii. Single nonrepeated inflationary injection

$$Y_t = (k_Y + \alpha)Y_{t-1} - (k_Y - \beta)Y_{t-2} + k_r r_M M \quad (3.10)$$

iv. Continuously repeated inflationary injection:

$$Y_t = (k_Y + \alpha)Y_{t-1} - (k_Y - \beta)Y_{t-2} + [1 + k_r r_M(t-1)]M \quad (3.11)$$

4. The Properties of the Integrated Model

The Particular Solutions: The permanent effects of the exogenous injections on the levels of income are given by the particular solutions of equations (3.8) to (3.11).

i. Single nonrepeated noninflationary injection: here the particular solution is obviously zero.

ii. Continuously repeated noninflationary injection: the particular solution is given by a constant z_1 , where

$$z_1 = \frac{1}{1 - \alpha - \beta} X > 0 \quad (4.1)$$

iii. Single nonrepeated inflationary injection: the particular solution is again given by a constant, z_2 , where

$$z_2 = \frac{k_r r_M}{1 - \alpha - \beta} M \quad (4.2)$$

Since $1 > k_r r_M > 0$, then it follows that $z_1 > z_2 > 0$, for $M = X$.

iv. Continuously repeated inflationary injection: the particular solution in this case is not given by a constant, but takes the form⁵ $z_3 + z_4 t$, where

$$z_3 + z_4 t = \frac{M}{1 - \alpha - \beta} \left[1 + \frac{k_r r_M [k_Y - (1 + \beta)]}{1 - \alpha - \beta} + k_r r_M \cdot t \right]. \quad (4.3)$$

Since $k_r r_M > 0$, income must expand divergently.

The Complementary Solution. In all four cases the reduced auxiliary equation is the same, viz.,

$$f(y) = y^2 - (k_Y + \alpha)y + (k_Y - \beta) = 0 \quad (4.4)$$

and from this the cyclical characteristics of the system can be easily determined. The roots of (4.4) can be separated immediately into the four following cases, since the modulus of the equation is $\sqrt{k_Y - \beta}$.

$$1. k_Y > [\beta + (1 + \sqrt{1 - \alpha - \beta})^2] \quad (4.5)$$

Real roots

$$2. [\beta + (1 + \sqrt{1 - \alpha - \beta})^2] > k_Y > (1 + \beta) \quad (4.6)$$

Complex roots with modulus greater than unity

5. For the calculation of this solution, which is little used in economic theory, see, e.g., W. L. Ferrar, "Higher Algebra," pp. 120-21.

$$3. (1 + \beta) > k_Y > [\beta + (1 - \sqrt{1 - \alpha - \beta})^2] \quad (4.7)$$

Complex roots with modulus less than unity

$$4. [\beta + (1 - \sqrt{1 - \alpha - \beta})^2] > k_Y \quad (4.8)$$

Real roots

Now

$$f(-1) = 1 + k_Y + \alpha + k_Y - \beta = 1 + 2k_Y + \alpha - \beta > 0 \quad (4.9)$$

$$f(0) = k_Y - \beta > 0 \quad (4.10)$$

$$f(+1) = 1 - k_Y - \alpha + k_Y - \beta = 1 - \alpha - \beta > 0 \quad (4.11)$$

Similarly, it can be seen that $f(y)$ is positive for any larger positive values of y and for any larger negative values of y . If, therefore, (4.4) has real roots *both* of them must lie within the ranges $(+\infty)$ to $(+1)$, $(+1)$ to 0 , 0 to (-1) and (-1) to $(-\infty)$, but since the sum of the roots is positive the negative roots are ruled out.

If $k_Y > [\beta + (1 + \sqrt{1 - \alpha - \beta})^2]$ it follows directly by squaring out the bracket that $k_Y > [(2 - \alpha) + 2\sqrt{1 - \alpha - \beta}]$. If this is so, then an analysis of the sum of the roots, $y_1 + y_2$, shows that

$$y_1 + y_2 = k_Y + \alpha > (2 - \alpha) + 2\sqrt{1 - \alpha - \beta} + \alpha = \\ 2(1 + \sqrt{1 - \alpha - \beta}) > 2. \quad (4.12)$$

Hence each of the roots must be greater than unity. Similarly, if $k_Y < [\beta + (1 - \sqrt{1 - \alpha - \beta})^2]$ then, by squaring out the bracket,

$$k_Y < [(2 - \alpha) - 2\sqrt{1 - \alpha - \beta}] \quad \text{and} \\ y_1 + y_2 = k_Y + \alpha < (2 - \alpha) - 2\sqrt{1 - \alpha - \beta} + \alpha = \\ 2(1 - \sqrt{1 - \alpha - \beta}) < 2 \quad (4.13)$$

so that each of the roots must be less than unity. Consequently, if (4.5) holds, each root of the auxiliary equation (4.4) is greater than unity, while if (4.8) holds, each root lies between unity and zero.

If (4.6) holds, it is obvious that cycles of ever-increasing amplitude are set up: with (4.7), cycles of ever-decreasing amplitude appear.

The General Solutions. Since the difference equations are of the simple linear type, the general solutions are obtained by adding the particular solutions, say Z , where Z takes one of the forms 0 or (4.1) to (4.3), to the complementary solution, say $A_1y_1^t + A_2y_2^t$, where y_1 and y_2 are the roots of the reduced auxiliary equation (4.4). Consequently,

$$Y_t = Z + A_1y_1^t + A_2y_2^t \quad (4.14)$$

where A_1 and A_2 depend on the initial conditions. The properties of this equation for all four forms of Z and the four possibilities for y_1 and y_2 are obvious, and are set out in the table on page 343.

THE ANATOMY OF MARKET FAILURE

By FRANCIS M. BATOR*

Introduction, 351. — I. The conditions of market efficiency, 353. — II. Neoclassical external economies: a digression, 356. — III. Statical externalities: an ordering, 363. — IV. Comments, 371. — V. Efficiency, markets and choice of institutions, 377.

What is it we mean by “market failure”? Typically, at least in allocation theory, we mean the failure of a more or less idealized system of price-market institutions to sustain “desirable” activities or to estop “undesirable” activities.¹ The desirability of an activity, in turn, is evaluated relative to the solution values of some explicit or implied maximum-welfare problem.

It is the central theorem of modern welfare economics that under certain strong assumptions about technology, tastes, and producers’ motivations, the equilibrium conditions which characterize a system of competitive markets will exactly correspond to the requirements of Paretian efficiency.² Further, if competitively imputed incomes are continuously redistributed in costless lump-sum fashion so as to achieve the income-distribution implied by a social welfare function, then the competitive market solution will correspond to the one electronically calculated Pareto-efficient solution which maximizes, subject only to tastes, technology and initial endowments, that particular welfare function.³

* Center for International Studies and Department of Economics, Massachusetts Institute of Technology.

I am much indebted to R. S. Eckaus and R. M. Solow for detailed comment and discussion.

1. “Activities” broadly defined, to cover consumption as well as production.

2. I.e., to the conditions which define the attainable frontier of maximal utility combinations with given preference functions, resource endowments and technology. A community is on its Paretian frontier if it is impossible to make anyone better off (in terms of his own ordinal preference function) without making someone else worse off. Associated with the utility possibility frontier, in turn, is a production possibility frontier denoting maximal alternative output combinations. (Cf. my “Simple Analytics of Welfare Maximization,” *American Economic Review*, XLVII (Mar. 1957), 22–59, and references therein.)

3. In other words, given the “right” lump-sum taxes, markets will match the allocation called for by the point of tangency of the relevant *W*-function with the utility-possibility frontier, i.e., by the “bliss point.” The *W*-function need not, of course, be explicit — it could be implicit in the political power-configuration which characterizes a community. On the other hand, it cannot be just any kind of function. It has to have some special characteristics which reflect a number of ethic-loaded restrictions, e.g., that individuals’ preference functions are to count, and to count positively (cf., *ibid.*, and Section V below).

Many things in the real world violate such correspondence: imperfect information, inertia and resistance to change, the infeasibility of costless lump-sum taxes, businessmen's desire for a "quiet life," uncertainty and inconsistent expectations, the vagaries of aggregate demand, etc. With most of these I am not here concerned: they have to do with the efficiency of "real life" market institutions operated by "real life" people in a nonstationary world of uncertainty, miscalculation, etc.

What follows is an attempt, rather, to explore and order those phenomena which cause even errorless profit- and preference-maximizing calculation in a stationary context of perfect (though limited) information and foresight to fail to sustain Pareto-efficient allocation. I am concerned, in other words, with the decentralizing efficiency of that regime of signals, rules and built-in sanctions which defines a price-market system.⁴

Specifically, Section I sets out the necessary conditions for efficiency of decentralized price-profit calculations both in a "laissez-faire" and in a "socialist" setting of Lange-Lerner civil servants. Section II is a brief digression on an often discussed mode of failure in these conditions: neoclassical external economies. It is concluded that the modern formulation of the doctrine, in terms of "direct interaction," begs more questions than it answers; further, that the usual emphasis on "divorce of scarcity from effective ownership" is misplaced. Section III, then, suggests a comprehensive ordering of types of market failure, with generalized indivisibility, public goods, and, last and not least, nonappropriability as the villains of the piece. Section IV consists of some comments on the Meade and Scitovsky classifications of external economies; on the analytical link between indivisibility and public goods; on the significance of "exclusion"; on organizational arrangements designed to offset externality; and on blends of the various types of market failure. Section V concludes with some cautionary notes on the relevance of market-efficiency for choice of institutions.

4. In most of what follows, I shall assume that individual preferences, though not necessarily sensitive only to own-consumption, are representable by strictly convex indifference surfaces (i.e., by an ordering (one for each individual) such that all points on a straight line connecting two equivalent points x and y are preferred to x (hence to y)). But convexity is too restrictive. It excludes not only such characteristics of man's psyche as violate the "usual" regularities — these I do want to exclude — but also such physical and topographical facts as lumpy consumption-goods. Rather than attempt a specification of preferences with convex-like properties where choice must be made among discrete bundles, I dodge the problem by attributing lumpiness only to inputs (including, however, inputs that are intermediate outputs).

I. THE CONDITIONS OF MARKET EFFICIENCY

The central theorem of modern welfare economics, the so-called *duality theorem*, asserts a correspondence between Pareto efficiency and market performance. Its analytical essence lies in the remarkable fact that with all-round convexity, independence of tastes, etc., the technocratically formulated, institutionally neutral, Paretian maximum-of-welfare problem has embedded within it a set of constants: "duals," Lagrangean multipliers, shadow-prices, which have all the analytical characteristics of prices, wages, rents, interest rates.⁵ Correspondence between Pareto-efficiency and market performance implies, at the least, that decentralized decisions in response to these "prices" by atomistic profit- and satisfaction-maximizers sustain just that constellation of inputs, outputs and commodity-distribution, that the maximum of the specified social welfare function calls for. It implies, in other words, that decentralized market calculations correctly account for all "economic" costs and benefits to which the relevant *W*-function is sensitive.⁶

Duality can fail in many ways. Specifically, and in a statical and "laissez-faire" context:⁷

(1) Duality will fail unless the Pareto-efficient (a) input-output points (production) and (b) associated commodity distribution points (exchange) which associate with the maximum of the welfare function in hand are characterized by a complete set of marginal-rate-of-substitution (*MRS*) equalities (or limiting inequalities) which, in

5. The theorem holds for the statical steady-state flow model of the Walrasian sort where the solution values are stationary time-rates; it holds, also, for dynamical systems involving capital formation (given, still, convexity throughout). For these last, the solution values are time paths of inputs, outputs, prices, etc. (A set of points is convex if, and only if, the straight lines connecting all possible pairs do not anywhere pass outside the set. The set of feasible output points bounded by a production possibility curve is convex, for instance, if the curve itself is concave-to-the-origin or a straight line. On all this, see Section V of "Simple Analytics," *ibid.*)

6. Given, again, optimal lump-sum redistribution of as-imputed incomes. While I make use of the lump-sum transfer device throughout this paper to abstract from the income distribution problem and permit exclusive attention to Pareto efficiency, it is well to note that this involves a measure of sleight-of-hand. No decentralized price-market type "game" can reveal the pattern of taxes and transfers that would maximize a particular welfare function. "Central" calculation — implicit if not explicit — is unavoidable. Moreover, since distribution (hence correct redistribution) of numeraire-incomes interdepends with allocation in production and exchange, the supposedly automatic, nonpolitical character of market mediation is a myth on the strictest neoclassical assumptions. This is not to say, even on our stratospheric levels of abstraction, that markets are "useless." Where they do compute well we are saved an awful lot of calculation.

7. With optimal redistribution.

turn, yield a set of price-like constants. Where no such constants exist, reference will be to *failure of existence*.⁸

(2) Should such an associated set of Lagrangean parameters exist, duality would nevertheless fail, specifically in production, unless the bliss configuration of inputs and outputs, evaluated in terms of these price parameters, will yield: (a) a local profit-maximum position for each producer, rather than, as possible, a profit minimum; (b) non-negative profits for all producers from whom production is required; (c) maximum profits-in-the-large for each producer. Failure on counts (a) and (c) will be labeled *failure by signal*, that on count (b) *failure by incentive*.⁹

(3) Even if all efficient production configurations, or the one which maximizes a particular welfare-function, coincide with points of maximum and non-negative producers' profits, market mediation may fail in production. If prices are determined by market forces, they will not correspond to a Paretian maximum unless self-policing perfect competition obtains in all markets. Self-policing competition requires "very many" producers in every market.¹ If, then, for whatever reason, some markets are saturated by a few firms of "efficient" scale, the full welfare-maximum solution of inputs, outputs and prices will not be sustained. There will be *failure by structure*.

(4) Finally, even if all above is satisfied, market performance could still fail, and fail in a statical sense, due to arbitrary legal and organizational "imperfections," or feasibility limitations on "keeping book," such as leave some inputs or outputs "hidden," or preclude their explicit allocation or capture by market processes (e.g., the restriction, unless I go into baseball, on the sale of the capitalized value of my lifetime services). Failure is *by enforcement*.

8. We could consider, instead, the configuration which associates with the initial pattern of ownership of endowment. Or we could play it safe and extend the conditions to cover each and every Pareto efficient configuration. But this would be overly strict, since many efficient situations have no relevance either to any interesting *W*-functions or in terms of the initial distribution of scarcities.

It may be worth noting, incidentally, that "existence," as used above, is not the same as existence in the sense of, e.g., Arrow and Debreu (in "Existence of an Equilibrium for a Competitive Economy," *Econometrica*, Vol. 22 (July 1954), pp. 265-90). They use the term to denote the complete set of conditions which defines competitive equilibrium, and this includes, in addition to all that is implied by (1) above, conditions akin to my conditions (2), and some analogous conditions on consumers.

9. This is slightly misleading: as we shall see, failure on count (c) leads both to signaling and to incentive troubles. Anyway, the labels are only for expository convenience.

1. Or at least the potentiality of very many producers, ready and able to "enter the fray" instantaneously. This may be sufficient in the constant-cost case, where the equilibrium number of firms per industry is indeterminate.

All the above are germane to duality in its usual sense, to the statical Pareto-efficiency of laissez-faire markets with genuine profit- and satisfaction-seekers.² Conditions (1), (2) and (4) are relevant, also, to the decentralizing efficiency of a Lange-Lerner type organizational scheme. In its "capitalist" version, with profit-motivated operation of privately-owned means of production where it is simply an anti-monopoly device to assure parametric take-prices-as-given behavior, conditions (1), (2) and (4) are all necessary for efficiency. Of course condition (3): self-policing competition, no longer matters.

In its true socialist version, a Lange-Lerner system can afford to "fail" also "by incentive," (2b). Socialist civil servants, under injunction to maximize profit (in the small) in terms of fixed centrally-quoted prices, care or should care not at all about absolute profitability. By assumption the scheme can dispense with the built-in incentive of positive profit: the lure of bureaucratic advancement, the image of Siberia, or the old school tie presumably substitute for the urge to get rich. But if prices and the injunction to maximize profit are to be used to decentralize, condition (1): existence, and (2a) and (2c): correct and unambiguous signals, remain crucial.³ So does condition (4): the solution of quantities and prices need not be profitable and self-enforcing, but it does have to be enforceable. If the nectar in apple blossoms is scarce and carries a positive shadow price, it must be possible to make every beekeeper pay for his charges' meals.

It warrants repetition that this has to do with whether a decentralized price-market game will or will not *sustain* a Pareto-efficient configuration. The word sustain is critical. There exists a host of further considerations which bear on dynamical questions of adjustment, of "how the system gets there." (E.g., will some "natural" price-market type computational routine of price-quantity responses with a meaningful institutional counterpart tend to track the solution?) These are not here at issue. We shall be concerned only with the prior problem of whether a price-market system which finds itself

2. The mathematically minded will object that (3) and (4), at least, do not really violate "duality" in its strict mathematical sense; the dual minimum problem still yields Lagrangean constants. True, yet I think it suggestive to use "duality" rather more loosely as a label for the general welfare theorem, particularly as this does not lead, in this context, to any ambiguity.

3. It is tempting, but wrong, to suggest that in a true Lange-Lerner world totals do not matter and only margins count. It is true that the non-negativeness of profits is immaterial. Where there is any sharing of shadow-price sets by two or more production points, however, totals necessarily become a part of the signaling system and if 2(c) does not hold they may lead down the garden path.

at the maximum-welfare point will or will not tend to remain there.⁴

The relevant literature is rich but confusing. It abounds in mutually reinforcing and overlapping descriptions and explanations of market failure: external economies, indivisibility, nonappropriability, direct interaction, public goods, atmosphere, etc. In a sense, our problem is simply to sort out the relations among these. In doing so, it is appropriate and useful to begin with a brief review of the neoclassical doctrine of external economies and of its modern formulation in terms of "direct interaction."

II. NEOCLASSICAL EXTERNAL ECONOMIES: A DIGRESSION

By Way of Some History

Marshall, as has often been pointed out, proposed the external economy argument to explain, without resort to dynamics, the phenomenon of a negatively sloped ("forward falling") long-run industry supply curve in terms consistent with a horizontal or rising marginal cost curve (MC) in the "representative" firm. The device permits — in logic, if not in fact — long-run competitive equilibrium of many firms within an industry, each producing at its profit-maximum price-equal-to-a-rising- MC position, without foreclosing the possibility of a falling supply price with rising industry output.⁵

The mechanism is simple. It is postulated that an expansion in the output of the industry as a whole brings into play economies which cause a downward shift of the cost curves of all the component

4. More precisely, whether the point of maximum welfare is or is not a point of self-policing and "enforceable" market equilibrium, where, following common usage, equilibrium is defined to subsume both the first-order and the second-order inequalities for a maximum. A firm, for instance, is taken to be in equilibrium only at a point of maximum profit. This way of defining equilibrium does bring in issues of stability, hence some implicit dynamics. In particular, the word "sustain" is taken to imply some scanning or reconnaissance by producers and consumers at least in the neighborhood of equilibrium. But I do not think it does any harm to subsume this much stability in the equilibrium notion. The possibility of a firm in *unstable* "equilibrium," i.e., in equilibrium at a point of minimum profit, is hardly likely to be of import.

On the other hand, correspondence between Pareto-efficiency and the equilibrium state of perfectly competitive markets is not sufficient to insure market efficiency. It is the burden of "failure by structure" that markets may fail to be competitive, and of "failure by enforcement" that legal or institutional constraints may prevent competitive markets from allocating efficiently, even though there does exist a competitive equilibrium for each Pareto-efficient configuration. "Existence" in the sense of Arrow and Debreu (*op. cit.*) is necessary but not sufficient for market-efficiency in the present context.

5. This refers to a so-called Marshallian supply curve. It has nothing whatever to do with the Walrasian "maximum quantity supplied at a given price" type schedule.

firms. These economies, however, are not subject to exploitation by any one of the myriad of tiny atomized firms. Their own MC curves, at $p = MC$, rise both before and after the shift, due, presumably, to internal diseconomies associated with the entrepreneurial function which defines the firm. Even the modern formulation is not entirely without ambiguity — institutional ambiguity is intrinsic to the device of parametrization: how many firms does it take for the demand curve of each to be perfectly horizontal? — but it does provide a means for “saving” the competitive model, of ducking the monopoly problem.

Marshall, and also Professor Pigou, “preferred,” as it were, the other horn of what they perhaps saw as a dilemma. The external economy device, while saving competition, implies a flaw in the efficacy of the “invisible hand” in guiding production.⁶ “Price equal to MC ” is saved, but wrong. Market forces, they argued, will not give enough output by industries enjoying external economies and will cause industries with rising supply curves to overexpand. Hence the Marshall-Pigou prescription: to harmonize private production decisions with public welfare, tax the latter set of industries and subsidize the former.

It took the better part of thirty years, and the cumulative powers of Allyn Young, and Messrs. Robertson, Knight, Sraffa, and Viner, to unravel the threads of truth and error which run through the Marshall-Pigou argument.⁷ The crucial distinction, which provides the key to it all, is between what Viner labeled technological external economies, on the one hand, and pecuniary external economies on the other. The latter, if dominant, cause the long-run supply curve of an industry, say A , to decline because the price of an input, B , falls in response to an increase in A ’s demand for it. The technological variety, on the other hand, though also a reversible function of industry output, consists in organizational or other improvements in efficiency which do not show up in input prices.⁸

6. That there are difficulties also with income distribution was by that time generally recognized.

7. The strategic articles, with the exception of Young’s (“Pigou’s *Wealth and Welfare*,” this *Journal*, XXVII (1913), 672–86), as well as Ellis and Fellner’s 1943 treatment, have all been reprinted in American Economic Association, *Readings in Price Theory*, ed. Stigler & Boulding. For an excellent modern discussion, see R. L. Bishop, *Economic Theory* (to appear).

8. Note, however, that there need be nothing about an organizational improvement to make it obvious in advance whether it will turn out to be technological or, through “internalization,” pecuniary. Many trade-association type services which are justified by the scale of an industry could as well be provided commercially, and vice versa.

As regards pecuniary external economies, Robertson and Sraffa made it clear that in a sense both the Marshall-Pigou conclusions were wrong. For one thing, no subsidy is called for. The implied gains in efficiency are adequately signaled by the input price, and profit-maximizing output levels by the A -firms are socially efficient. Second, monopoly troubles may be with us, via, as it were, the back door. For what causes the price of B to drop in response to increased demand? We are back where we started: a declining long-run supply curve.

In the end, then, if *internal* technological economies of scale are ruled out, we are left with only *technological* external economies. All pecuniary external economies must be due to technological economies somewhere in the system.⁹ It is true — and this is what remains of the original Marshall-Pigou proposition — that technological externalities are not correctly accounted for by prices, that they violate the efficiency of decentralized market calculation.

*The Modern Formulation*¹

In its modern version, the notion of external economies — external economies proper that is: Viner's technological variety — belongs to a more general doctrine of "direct interaction." Such interaction, whether it involves producer-producer, consumer-consumer, producer-consumer, or employer-employee relations, consists in interdependences that are external to the price system, hence unaccounted for by market valuations. Analytically, it implies the nonindependence of various preference and production functions. Its effect is to cause divergence between private and social cost-benefit calculation.

That this is so, is easily demonstrated by means of a simplified variant of a production model suggested by J. E. Meade.² Assume a world of all-round perfect competition where a single purchasable and inelastically supplied input, labor (\bar{L}), is used to produce two homogeneous and divisible goods, apples (A) and honey (H), at nonincreasing returns to scale. But while the output of A is dependent only on L_A : $A = A(L_A)$, honey production is sensitive also to the level of apple output: $H = H(L_H, A(L_A))$. (Professor Meade

9. Pecuniary diseconomies, in contrast, need have no technological counterpart. Finite-elastic supplies of unproduced inputs are a sufficient cause. Recall, incidentally, that only narrowly statical reversible phenomena are admissible here.

1. While this section makes some slight use of elementary calculus, the reader uninterested in technicalities may avoid, without loss of continuity, all but some simple notation.

2. *Economic Journal*, LXII (Mar. 1952). Meade uses a two factor model and, while he does not explicitly solve the Paretian maximum problem, shows that market imputed rates of remuneration will not match marginal social product.

makes pleasurable the thought of apple blossoms making for honey abundance.)³

By solving the usual constrained maximum problem for the production-possibility curve, it can be shown that Paretian production efficiency implies

$$p_H \frac{\partial H}{\partial L_H} = w \quad (1)$$

$$p_A \frac{dA}{dL_A} + p_H \frac{\partial H}{\partial A} \frac{dA}{dL_A} = w \quad (2)$$

where p_H , p_A , and w represent the prices, respectively, of honey, apples and labor.⁴ Equation (1) is familiar enough and consistent with profit maximizing. Each competitive honey producer will do for profit what he must for efficiency: hire labor until the value of its social as well as private marginal product equals the wage rate. Not

so the apple producers; unless $\frac{\partial H}{\partial A} = 0$ — unless the cross effect of

apples on honey is zero — their profit-maximizing production decisions will be nonefficient. Specifically, if apples have a positive external effect on honey output, market-determined L_A will be less than is socially desirable.⁵

A different way to see this is to examine the relations of private to social marginal cost. The marginal money cost of apples to the com-

3. Both functions are assumed homogeneous of degree one. Moreover, apple blossoms (or the nectar therein) are exhaustible, rationable "private" goods: more nectar to one bee means less to another. On the need for this assumption, see Section III-3 below.

4. Assuming internal tangencies and all-round convexity (the last is implicit in constant returns to L : the A -effect on H reinforces convexity), as well as non-satiation and nonredundancy ($\bar{L} = L_A + L_H$), the maximization of $p_A A + p_H H$, subject to the production functions and the supply of labor, is equivalent to finding a critical value for the Lagrangean expression, $F = p_A A(L_A) + p_H H[L_H; A(L_A)] + w(\bar{L} - L_A - L_H)$. To do so, differentiate F with respect to L_A and L_H , treating p_A , p_H and w as arbitrary constants and set the resulting first order partial derivatives equal to zero. This will give exactly (1) and (2). (Needless to say, the value weights can be varied at will, or taken as given.)

5. To see this, rewrite (2) to read $\frac{dA}{dL_A} = \frac{w}{p_A + p_H \frac{\partial H}{\partial A}}$ and match it against

the profit-maximizing rule, $\frac{dA}{dL_A} = \frac{w}{p_A}$. Clearly, $\frac{\partial H}{\partial A} \leq 0 \rightarrow$

$$\left(\frac{dA}{dL_A} \right)_{\text{Private}} \leq \left(\frac{dA}{dL_A} \right)_{\text{Social}}$$

petitive apple producer is $\frac{w}{dA/dL_A}$; that of honey to the beekeeper,

$\frac{w}{\partial H/\partial L_H}$. It is the ratio of the two: $\frac{\partial H/\partial L_H}{dA/dL_A}$, that competitive market-

mediation brings into equality with the equilibrating configuration of relative prices. Markets will be efficient if, and only if, this *private* marginal cost ratio reflects the true marginal cost to society of an extra apple in terms of foregone honey: the marginal rate of transformation between H and A .

What is *MRT* in the model? Differentiating (totally) the two production functions and dividing the value of one derivative into the other, we get, in absolute (cost) terms:

$$MRT \equiv \left| \frac{dH}{dA} \right| = \frac{\partial H/\partial L_H}{dA/dL_A} - \frac{\partial H}{\partial A}.$$

If, then, $\frac{\partial H}{\partial A} > 0$, the true marginal *social* cost of an "extra"

apple, in terms of honey foregone, is less than the market-indicated private cost. It is less precisely by the amount of positive "feed-back" on honey output due the "extra" apple.

By combining (1) and (2), eliminating w , and dividing through by p_H and $\frac{dA}{dL_A}$, we get the condition for Pareto efficiency in terms of private *MC*'s:

$$\frac{\partial H/\partial L_H}{dA/dL_A} = \frac{p_A}{p_H} + \frac{\partial H}{\partial A}.$$

Clearly, price equal to private marginal cost will not do. Further, if prices are market-determined, they will diverge from true, *social* marginal cost.

Any number of variations on the model suggest themselves. As Meade pointed out, interactions can be mutual and need not be associated with the outputs. Even in the above case, it is perhaps more suggestive to think of L_A as producing some social value-product both in the A industry and the H industry. In the most general formulation, one can simply think of each production function as containing all the other variables of the system, some perhaps with zero weight. Moreover, by introducing two or more nonproduced inputs one can, as Meade does, work out the consequences for income distribution and input proportions.⁶

6. The question of whether technological external economies involve shifts of each other's production functions, or mutually induced movements along such functions, is purely definitional. If one chooses so to define each producer's

Some Queries

The modern formulation of the doctrine of external economies, in terms of direct interaction, is not only internally consistent: it also yields insight. Yet one may well retain about it some dissatisfaction. There is no doubt that the Robertson-Sraffa-Viner distinction between the technological and the pecuniary sort gets to the nub of what is the matter with the original Marshallian analysis. It cuts right through the confusion which led Marshall and Pigou to conclude that the price mechanism is faulty in situations where in truth it is at its best: in allocating inputs in less than infinitely elastic supply between alternative productive uses. It also facilitates unambiguous formulation of the more difficult "falling supply price" case. But in a sense it only begs the fundamental question: what is it that gives rise to "direct interaction," to short circuit, as it were, of the signaling system?

Most modern writers have let matters rest with the Ellis-Fellner type explanation: "the divorce of scarcity from effective ownership."⁷ Does nonappropriability then explain all direct interaction? In a sense it does, yet by directing attention to institutional and feasibility considerations which make it impracticable for "real life" market-institutions to mimic a price-profit-preference computation, it diverts attention from some deeper issues. Surely the word "ownership" serves to illuminate but poorly the phenomenon of a temperance leaguer's reaction to a hard-drinking neighbor's (sound insulated and solitary) Saturday night, or the reason why a price system, if efficient, will not permit full "compensation," in an age of electronic scramblers, for an advertisement-less radio program, or for the "services" of a bridge.⁸

function as to give axes only to inputs and outputs that are purchased and sold, or at least "controlled," and the effects of everything else impinging on production (e.g., of humidity, apple blossoms, etc.) are built into the curvature of the function, then it follows that externalities will consist in shifts of some functions in response to movements along others. On the other hand, if, as in our apple-honey case, it seems useful to think of the production function for H as having an A -axis, then, clearly, induced movement along the function is a signal of externality.

7. *Op. cit.*

8. Moreover, in the one sense in which nonappropriability fits all cases of direct interaction, it explains none. If all it denotes is the failure of a price-market game properly to account for (to appropriate) all relevant costs and benefits, then it is simply a synonym for market failure (for generalized externality), and cannot be used to explain what causes any particular instance of such failure. I use it in a much narrower sense, to mean the inability of a producer of a good or service physically to exclude users, or to control the rationing of his produce among them. In my sense not only bridges but also, say, television programs are fully appropriable: it is always possible to use scramblers.

It may be argued, of course, that at least the two latter examples are out of order, that radio programs and bridges do not involve "direct," i.e., non-price, interaction. But is this really so? Does not the introduction of a new program directly affect my and your consumption possibilities, in ways other than by a change in relative prices? Does not a bridge, or a road, have a direct effect on the production possibilities of neighboring producers, in precisely the sense in which apples affect the possibilities of beekeepers?⁹

True, perhaps bridges and roads are unfair: they violate the neo-classical assumption of perfect divisibility and nonincreasing returns to scale. But they surely do involve non-price interaction. In fact, lumpiness and increasing returns are perhaps the most important causes of such interaction. Are they to be denied status as externalities? More generally, are we to exclude from the class of externalities any direct interaction not due to difficulties with "effective ownership," any failures other than "by enforcement"?

It would be, of course, perfectly legitimate to do so — tastes are various. But I think it more natural and useful to broaden rather than restrict, to let "externality" denote any situation where some Paretian costs and benefits remain *external* to decentralized cost-revenue calculations in terms of prices.¹ If, however, we do so, then clearly nonappropriability² will not do as a complete explanation. Its concern with the inability of decentralized markets to sustain the solution-prices and quantities called for by a price-profit-preference type calculation, as computed by a team of mathematicians working with IBM machines, tends to mask the possibility that such machine-

9. It is possible, of course, to interpret these examples as involving very large changes in price: from infinity to zero. But it does not help to do so. The shared characteristic of bridges and programs is that there is no price which will efficiently mediate both supply and demand.

I have puzzled over ways of limiting the notion of "direct interaction" to something less than all instances where there is some interaction not adequately signaled by price. Robert Solow has suggested to me that this might be done by distinguishing situations where something is not subject to a market test at all from instances where no single price constitutes a correct test for both sides of a transaction (e.g., where the correct ration price for the services of an expensive facility is zero). I am inclined, rather, to drop the attempt to use "direct interaction" as an explanation of market failure; it is best used, if at all, as yet another synonym for such failure.

1. Recall that it is the existence of such "externality," of residue, at the bliss-point, of Pigouvian "uncompensated services" and "incidental uncharged disservices" that defines market failure. It may be objected that to generalize the externality notion in this way is to rob it of all but descriptive significance. But surely there is not much to rob; even in its strictest neoclassical formulation it begs more than it answers. In its generalized sense it at least has the virtue of suggesting the right questions.

2. As defined in fn. 8, p. 361 above.

calculated solution q 's may well be nonefficient.³ It explains failure "by enforcement," but leaves hidden the empirically more important phenomena which cause failure by "nonexistence," "signal," and "incentive." Section III is designed to bring these deeper causes of generalized externality into the foreground.

III. STATICAL EXTERNALITIES: AN ORDERING

If nonappropriability is, by itself, too flimsy a base for a doctrine of generalized (statical) externality, what broader foundation is there? Section I's hierarchy of possible modes of market failure suggests a fivefold classification. If, however, one looks for an organizing principle not to modes of failure but to causes, there appear to be three polar types: (1) Ownership Externalities, (2) Technical Externalities,⁴ and (3) Public Good Externalities. These are not mutually exclusive: most externality phenomena are in fact blends. Yet there emerges a sufficient three-cornered clustering to warrant consolidation.⁵

Type (1): Ownership Externalities

Imagine a world which exhibits generalized technological and taste convexity, where the electronically calculated solution of a Paretian maximum-of-welfare problem yields not only a unique set of inputs, outputs and commodity-distribution, but where initial endowments plus lump-sum transfers render income distribution optimal in terms of the community's social welfare function. Assume, further, that everything that matters is divisible, conventionally rationable, and either available in inelastic total supply,⁶ or producible at constant returns to scale; also that tastes are sensitive only to own-consumption. We know, then, from the duality theorem, that

3. Or that the algorithm may break down for lack of a consistent set of p 's.

4. I should much prefer "technological," but since this would necessarily confuse my Type (2) with Professor Viner's "technological" I fixed on "technical."

5. In effect, we end up with a five-by-three ordering of types of "failure": five "modes" vs. three "causes." Its relation to Meade's categories (*op. cit.*) and to Tibor Scitovsky's classification (in "Two Concepts of External Economies," *Journal of Political Economy*, LXII, April 1954) is discussed in Section IV below. I have had the benefit of reading, also, William Fellner's "Individual Investment Projects in Growing Economies," *Investment Criteria and Economic Growth* (Proceedings of a Conference, Center for International Studies, Massachusetts Institute of Technology, 1955) and an unpublished paper by Svend Laursen, "External Economies and Economic Growth."

6. The supply of such nonproduced scarcities need not, of course, remain constant. On the other hand, their ownership distribution must not be so concentrated as to preclude competitive rationing. There must exist no "indivisible" lake full of fish, etc., such as might be subject to monopolization, but thousands of lakes, all perfect substitutes.

the bliss point implies a unique⁷ set of prices, wages and rents, such as would cause atomistic profit- and preference-maximizers to do exactly what is necessary for bliss. In particular, all required production points give maximum and non-negative producer's profits.

This is an Adam Smith dream world. Yet it is possible that due to more or less arbitrary and accidental circumstances of institutions, laws, customs, or feasibility, competitive markets would not be Pareto-efficient. Take, for instance, the Meade example of apples and honey. Apple blossoms are "produced" at constant returns to scale and are (we assumed) an ordinary, private, exhaustible good: the more nectar for one bee, the less for another. It is easy to show that if apple blossoms have a positive effect on honey production (and abstracting from possible satiation and redundancy) a maximum-of-welfare solution, or any Pareto-efficient solution, will associate with apple blossoms a positive Lagrangean shadow-price.⁸ If, then, apple producers are unable to protect their equity in apple-nectar and markets do not impute to apple blossoms their correct shadow value, profit-maximizing decisions will fail correctly to allocate resources (e.g., L) at the margin. There will be failure "by enforcement."

This is what I would call an *ownership* externality. It is essentially Meade's "unpaid factor" case. Nonappropriation, divorce of scarcity from effective ownership, is *the* binding consideration. Certain "goods" (or "bads") with determinate non-zero shadow-values are simply not attributed. It is irrelevant here whether this is because the lake where people fish happens to be in the public domain, or because "keeping book" on who produces, and who gets what, may be impossible, clumsy, or costly in terms of resources.⁹ For whatever legal or feasibility reasons, certain variables which have positive or negative shadow value are not "assigned" axes. The beekeeper thinks only in terms of labor, the orchard-owner only in terms of apples.

The important point is that the difficulties reside in institutional arrangements, the feasibility of keeping tab, etc. The scarcities at issue are rationable and finely divisible and there are no difficulties with "total conditions": at the bliss-configuration every activity would pay for itself. Apple nectar has a positive shadow price, which

7. Or, where there are corners, only inessentially indeterminate.

8. Set up a variant of the Apple-Honey model of Part II, introducing apple blossoms, B , explicitly. Add a production function, $B = B(L_A)$, and substitute $B(L_A)$ for $A(L_A)$ as the second input in honey production. The solution will give out a positive Lagrangean shadow price for B , and profit-maximizing producers of the joint products: A and B , will push L_A to the socially desirable margin.

9. Though on this last, see Section IV, first paragraph.

would, if only payment were enforceable, cause nectar production in precisely the right amount and even distribution would be correctly rationed. The difficulty is due exclusively to the difficulty of keeping accounts on the nectar-take of Capulet bees as against Montague bees.¹

Many of the few examples of interproducer external economies of the reversible technological variety are of this type: "shared deposits" of fish, water, etc.² Much more important, so are certain irreversible dynamical examples associated with investment. For instance, many of Pigou's first category of externalities: those that arise in connection with owner-tenant relationships where durable investments are involved, have a primarily organizational quality.³ Perhaps the most important instance is the training of nonslave labor to skills — as distinct from education in a broader sense (which partakes more of Type (3)). In the end, however, and in particular if restricted to reversible statical cases, it is not easy to think of many significant "ownership externalities" pure and simple. Yet it turns out that only this type of externality is really due to nonappropriability.

Type (2): Technical Externalities

Assume, again, that all goods and services are rationable, exhaustible, scarcities, that individual ordinal indifference maps are convex and sensitive only to own-consumption and that there exist no ownership "defects" of Type (1). If, then, the technology exhibits indivisibility or smooth increasing returns to scale in the relevant range of output, these give rise to a second and much more important type of market failure: "technical externality."⁴

1. More generally, it could as well be due to difficulty in knowing who "produced" the "benefit" — oil wells drawing on the same pool are an example. The owner cannot protect his own; in fact it is difficult to know what one means by "his own." Moreover, in the case of *diseconomies*, at least, it may be that both the source and the recipient of the "bad" are identified: one factory producing soot and nothing but one laundry in the neighborhood, yet it is difficult to see how a price can be brought to bear on the situation. Presumably the laundry can pay for negative units of smoke.

2. Though indivisibility elements enter into some of these. Why can't somebody "own" part of a lakeful of fish?

3. When not simply due, in a world of uncertainty, to inconsistent expectations.

4. Again, this is not the same as Viner's "technological." Note, incidentally, that the above formulation unabashedly begs the question of whether smooth increasing returns to scale could or could not arise without indivisibility somewhere. The issue is entirely definitional: it is conceptually impossible to disprove either view by reference to empirical evidence. (Cf. "Simple Analytics," *loc. cit.*, fn. 37 and references.) (Continued on page 366.)

The essential analytical consequence of indivisibility,⁵ whether in inputs, outputs or processes, as well as of smooth increasing returns to scale, is to render the set of feasible points in production (input-output space) nonconvex. A connecting straight line between some pairs of feasible points will pass outside the feasible set. Nonconvexity, in turn, has a devastating effect on duality.⁶

In situations of pure "technical externality" there does, of course, still exist a maximal production possibility frontier (*FF*); and with a Samuelson-type social indifference map (*SS*) — i.e., a map "corrected" for income distribution which provides a ranking for the community as a whole of all conceivable output combinations⁷ — it is possible, in concept, to define a bliss point(s).⁸ Also, where indivisibility is exhibited by outputs, and only outputs, or, stronger, where smoothly increasing returns to scale is the only variety of nonconvexity — isoquants for one, are properly convex — the locus of efficient output combinations can be defined in terms of conditions on marginal-rates-of-input-substitution.⁹ Moreover, bliss could possibly occur at a point where *SS* is internally tangent to *FF*, perhaps to a convex *FF*. But even in the least "pathological," most neoclassi-

The pioneer work on decreasing cost situations is Jules Dupuit's remarkable 1844 essay, "On the Measurement of Utility of Public Works," translated in *International Economic Papers*, No. 2, ed. A. T. Peacock, *et al.* Harold Hotelling's "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," in the July 1938 issue of *Econometrica*, is the originating modern formulation. Cf., also, references to work by R. Frisch, J. E. Meade, W. A. Lewis and others in Nancy Ruggles' excellent survey articles on marginal cost pricing (*Review of Economic Studies*, XVII (1949-50), 29-46, and 107-26).

5. Indivisibility means lumpiness "in scale" and not the kind of indivisibility-in-time we call durability. (Durability, as such, does not violate convexity.) Lumpiness has to do with the impossibility to vary continuously, e.g., the capacity service-yield per unit time of such things as bridges.

6. The best known and perhaps most important variety of nonconvexity occurs where isoquants are properly convex, but returns to scale are increasing, hence the full set of feasible input-output points is nonconvex. (In a two-input, one-output situation, slices by (vertical) planes through the origin perpendicular to the input plane will cut the production surface in such a way as to give a nonconvex boundary.) A production point lying in an "increasing returns" region of a production function implies that (1) the associated average cost curve (*AC*) is downward sloping at that level of output; (2) the associated marginal cost curve (*MC*), while it may be rising, could as well be falling and will certainly lie below *AC*; and (3) the production possibility curve of the community may be nonconvex. On all this, see Part V of "Simple Analytics," *loc. cit.*

7. Cf. P. A. Samuelson, "Social Indifference Curves," this *Journal*, LXX (Feb. 1956), 1-22. Such a function presumes that *numeraire*-incomes are continuously redistributed so as to maximize in utility space over the community's operative social welfare function.

8. This is saying very little, of course, except on the level of metaphysics.

9. Inequalities due to kinks and corners are as good as equalities where all is smooth.

cally well-behaved case, where there exists a meaningfully defined set of shadow prices associated with the bliss point, genuinely profit-seeking competitive producers, responding to that set of prices, would fail to sustain optimal production. At best, even if at the bliss-configuration all MC 's are rising, some producers would have to make continuing losses, hence would go out of business; market calculations would necessarily fail "by incentive." If, in turn, prices are not centrally quoted but permitted to set themselves, monopoly behavior will result. There will be failure "by structure."

Further, bliss may require production at levels of output where losses are not only positive, but at a constrained maximum;¹ $p = MC$ may be correct, though MC at that point is falling. If so, the embedded Lagrangean constants may still retain meaning as marginal rates of transformation, but they will fail to sustain efficient production even by Lange-Lerner civil servants who care only about margins and not about absolute totals. There will be failure "by signal": producers under injunction to maximize profit (in the small) will not remain where they ought to be.

If, moreover, we drop the assumption of smooth increasing returns to scale and permit indivisibilities such as give scallop-like effects and kinks in cost curves and in the production-possibility curve, things get even more complicated. Bliss could require production at points of positive but locally minimum profit, where MC exceeds AC but is falling. Worse, even if bliss should occur at points where production functions are locally convex and MC (greater than AC) is rising, prequoted prices may still not sustain the solution unless production functions are in fact convex throughout. Though positive and at a local maximum, profits may not be at their maximum-maximorum: other hills with higher peaks may induce producers with vision at a distance to rush away from bliss. Alternatively, if prices are not administered, competition may not be self-policing and markets could fail "by structure."²

1. Subject to the requirement that total cost for that level of output be a minimum, i.e., that each producer be on his least-cost expansion path.

2. Where sharp indivisibility gives a nonconvex production possibility curve with corners and kinks, duality may fail even if there exists a price vector in terms of which decentralized producer-calculations would sustain the bliss-point output mix. The existence of such a vector does not assure that it will coincide with the price-vector which would efficiently ration that bill of goods among consumers. The point is that there may not exist a *single* set of prices which will at the same time keep both consumers and producers from rushing away from where they ought to be. The prices which will effectively mediate production may cause consumers' calculations to go wrong and vice versa.

It should be noted, incidentally, that none of the above takes space and distance considerations into account. For some interesting effects of plant-indivisi-

On the other hand, given our assumptions, the Paretian contract locus of maximal (ordinal) utility combinations which is associated with any one particular output point is defined, as in the trouble-free neoclassical model, by the usual subjective, taste-determined, marginal-rate-of-substitution equalities (or, at corners, inequalities). These *MRS* equalities, in turn, imply a set of shadow-prices which, if centrally quoted, would efficiently ration among consumers the associated (fixed) totals of goods. In the sphere of exchange, then, a decentralized price system works without flaw.

In what sense do these Type (2) situations exhibit "externality"? In the (generalized) sense that some social costs and benefits remain external to decentralized profitability calculations. With Type (1) externalities, though it is not feasible to police the bliss values of all quantities and prices, there exists embedded in the solution a set of prices whose use for purposes of decentralized signaling would sustain, if only appropriation or exclusion were feasible, both itself and the maximum welfare configuration of inputs, outputs, and distribution. This is not the case here. In Type (1) situations, at the bliss point there is complete correspondence between social and private pay-off, both at the margin and in totals.³ Profits are at their maxima and non-negative throughout. Here there is no such correspondence; there may well be divergence, either at the margin: bliss-profits may be at a "minimum," or in *totals*. The private totals in terms of which producers in an (idealized) market calculate — total revenue minus total cost — will not reliably signal the social costs and benefits implied by the relevant social indifference curves.⁴ Hence at the set of prices which would correctly ration the bliss point bill of goods, that bill of goods may not be produced by profit seekers, or even by Lange-Lerner civil servants.⁵

bility where there are interplant flows and transport takes resources, see T. C. Koopmans and M. Beckmann, "Assignment Problems and the Location of Economic Activities," *Econometrica*, Vol. 25 (Jan. 1957).

3. More correctly, there would be such correspondence, if only the p 's could be policed.

4. This is particularly awkward since the very nonconvexities which cause a divergence between private and social total conditions render output-mix calculations based on margins alone wholly inadequate. Even if bliss gives all local profit maxima, there may be several such open to any one producer, hence he must make total calculations in order to choose.

5. There is one qualification to be made to the above. It may be that the bliss configuration gives unique and positive profit maxima throughout, though some production functions exhibit nonconvexities at a distance. It was to exclude this case that we assumed that increasing returns or indivisibility obtain in the "relevant ranges." Should this happen, no "externality" divergence of social and private calculation will occur, at least in a statical context. But unless all

A point to note, in all this, is that in relation to "technical externalities" the nonappropriability notion, as generally conceived, tends to miss the point. Strictly speaking, it is, of course, true that price mediation, if efficient, cannot be counted on to "appropriate" the full social benefits of activities showing increasing returns to scale or other types of indivisibility to those engaged in them. But the existence of such "uncompensated services" has in this case nothing whatever to do with "divorce of scarcity from ownership," with feasibility limitations on "exclusion." It is entirely feasible to own a bridge and profitably ration crossings; indeed, a private owner would do so. The point is, rather, that such profitable rationing, such "compensation" for services rendered, would inefficiently misallocate the "output" of bridge crossings. If in terms of scarce resource inputs the marginal cost of an additional crossing is zero, any positive toll will, in general, have the usual monopolistic effect: the resulting output configuration will not be efficient.⁶

This, incidentally, is where most pecuniary external economies lead: a supplier is required to produce in a range of declining AC due to internal technological economies of scale and hence cannot make "ends meet" at the socially correct price. The crucial associated difficulty at the level of social organization is monopoly.

Can we leave matters at that? Not quite. There is a third kind of externality, recently emphasized by Professor Samuelson, caused by so-called "public goods."

Type (3): Public Good Externalities

In some recent writings on public expenditure theory, Samuelson has reintroduced the notion of the collective or public good. The defining quality of a pure public good is that "each individual's consumption of such a good leads to no subtractions from any other individual's consumption of that good . . .",⁷ hence, "it differs from a private consumption good in that each man's consumption of it, X_1^1 and X_2^2 respectively, is related to the total X_2 by a condition of

is convex throughout, the existence of such a locally stable tangency cannot be taken as evidence that the point is in fact the bliss-point — a difficulty of considerable significance for dynamical efficiency.

6. Of course, if at bliss the bridge were to be used "to capacity," it is possible that the Lagrangean ration price (now positive) would make commercial operation profitable. If so, an administered price setup would efficiently mediate the demand and supply of crossings. But while a Lange-Lerner system would work fine, laissez-faire markets would fail "by structure."

7. P. A. Samuelson, *Review of Economics and Statistics*, XXXVI (Nov. 1954), 387.

equality rather than of summation. Thus, by definition, $X_2^1 = X_2$ and $X_2^2 = X_2$.⁸

As Samuelson has shown, the form of the marginal rate of substitution conditions which define the Pareto-efficient utility possibility frontier in a world where such public goods exist, or at least where there are outputs with important "public" qualities, renders any kind of price-market routine virtually useless for the computation of output-mix and of distribution, hence, also, for organizational decentralization. Where some restraints in the maximum problem take the form: total production of X equals consumption by Crusoe of X equals consumption of X by Friday, Pareto efficiency requires that the marginal rate of transformation in production between X and Y equal not the (equalized) MRS of each separate consumer, but rather the algebraic sum of such MRS 's. This holds, of course, in what in other respects is a conventionally neoclassical world: preference and production functions are of well-behaved curvature, all is convex.

If, then, at the bliss point, with Y as numeraire, P_x is equated to the marginal Y -cost of X in production (as is required to get optimal production), and X is offered for sale at that p_x , preference-maximizing consumers adjusting their purchases so as to equate their individual MRS 's to p_x will necessarily under-use X . Moreover, a pricing game will not induce consumers truthfully to reveal their preferences. It pays each consumer to understate his desire for X relative to Y , since his enjoyment of X is a function only of total X , rather than, as is true of a pure private good, just of that fraction of X he pays for.

The two Samuelson articles⁹ explore both the analytics and the general implications of "public goods." Here the notion is of relevance because much externality is due precisely to the "public" qualities of a great many activities. For example, the externality associated with the generation of ideas, knowledge, etc., is due in good part to the public character of these "commodities." Many interconsumer externalities are of this sort: my party is my neighbor's disturbance, your nice garden is any passerby's nice view, my children's education is your children's good company, my Strategic Air Command is your Strategic Air Command, etc. The same consumption item enters, positively or negatively, both our preference func-

8. P. A. Samuelson, *Review of Economics and Statistics*, XXXVII (Nov. 1955), 350.

9. And a third unpublished paper, which was read at the 1955 American Economic Association meetings and to a copy of which I came to have access while this paper was being written. For earlier writings on public goods, by Wicksell, Lindahl, Musgrave, Bowen and others see references in the above cited Samuelson articles.

tions. The consumptions involved are intrinsically and essentially joint.

This kind of externality is distinct from either of the other two pure types. Here technological nonconvexities need in no way be involved. In fact the $MRT = \Sigma MRS$ condition is certain to hold true precisely where production takes place at constant or non-increasing returns, and hence where the production possibility set is necessarily convex. Further, there are no decentralized organizational rearrangements, no private bookkeeping devices, which would, if only feasibility were not at issue, eliminate the difficulty. It is the central implication of the Samuelson model that where public good phenomena are present, there does not exist a set of prices associated with the (perfectly definable) bliss point, which would sustain the bliss configuration. The set of prices which would induce profit-seeking competitors to produce the optimal bill of goods, would be necessarily inefficient in allocating that bill of goods. Moreover, even abstracting from production, no single set of relative prices will efficiently ration any fixed bill of goods so as to place the system on its contract locus, except in the singular case where at that output and income-distribution MRS 's of every individual are identically the same (or zero for all but one). There is failure "by existence."

IV. COMMENTS

Type (1). In a sense, Type (1) is not symmetrical with the other two categories. One can think of some nontrivial instances where the institutional element does appear to be "binding": skill-training of people, for example. But even there, it could be argued that the crucial elements are durability, uncertainty, and the fact that slavery as a mode of organization is itself in the nature of a public good which enters people's preference functions, or the implicit social welfare function, inseparably from the narrowly "economic" variables. In those instances, in turn, where bookkeeping feasibility appears to be the cause of the trouble, the question arises why bookkeeping is less feasible than where it is in fact being done. In the end, it may be that much of what appears to partake of Type (1) is really a compound of Types (2) and (3), with dynamical durability and uncertainty elements thrown in. At any rate, a deeper analysis of this category may cause it substantially to shrink.

Nonproduced scarcities. One particular instance where what appears like Type (1) is really Type (2) warrants special mention. Public ownership of nonproduced resources, e.g., the lakes and mountains of national parks, may make it appear that externality is due

to statutory barriers to private ownership and commercial rental. But this is missing the point. Take, for instance, a community which has available a single source of fresh water of fixed capacity. Assume that the bliss solution gives out a positive ration-price per gallon such as would make sale of the water commercially profitable. Yet a laissez-faire system would fail, "by structure," to sustain bliss. A private owner of the single indivisible well, if given his head, would take advantage of the tilt in the demand curve. The real cause of externality is not the arbitrary rapaciousness of public authority but the indivisibility of the source of supply. This case, by the way, is akin to where indivisibility or increasing returns to scale within a range allow profitable scope for one or a few efficient producers, but for no more. At the bliss price all will do the right thing, but if prices are not administered, oligopoly or monopoly will result. A capitalist Lange-Lerner system with private ownership but administered prices would work fine, but laissez-faire markets would fail.

Meade's "atmosphere." The relation of my tri-cornered ordering to Meade's polar categories is of interest.¹ His first category, "unpaid factors," is identical to my Type (1). But his second, labeled "atmosphere," is a rather curious composite. Meade's qualitative characterization of "atmosphere": e.g., of afforestation-induced rainfall, comes very close to the public good notion.² He links this, however, as necessarily bound up with increasing returns to scale in production to society at large, hence a J. B. Clark-like overexhaustion, adding-up problem.³

If, following Meade, one abstracts from shared water-table phenomena (let rain-caused water input be rigidly proportional to area) then Farmer Jones' rain is Farmer Smith's rain and we have my Type (3). But nothing in this situation requires that either farmer's full production function (with an axis for rain) need show increasing returns to scale. It may be that returns to additional bundles of non-rain inputs, with given constant rainfall, diminish sharply, and that it takes proportional increases of land, labor *and* rain to get a proportional effect on output. If so, Meade's overexhaustion problem

1. *Op. cit.* (This and the next section can be omitted without loss of continuity.)

2. See esp. bottom of p. 61 and top of p. 62, *op. cit.*

3. Since his argument is restricted to competitive situations, hence necessarily excludes increasing-returns-to-paid-factors such as would require production at a loss, Meade specifies constant returns to proportional variation of labor and land in wheat farming, though the full production function for wheat, including the atmosphere input (rain), exhibits increasing returns to scale. But the individual farmer does not pay for rain, hence his factor payments just match his sales revenue, by the Euler Theorem.

will not arise. But all would not be well: the public good quality of rainfall would cause an independent difficulty, one that Meade, if I understand him correctly, does not take into account, i.e., that rain ought to be "produced" by timber growers until its MC is equal to the sum of all the affected farmers MRS 's for rain as an input, whatever may be the curvature of the latter's production functions.⁴

On the other hand, Meade's formal mathematical treatment of "atmosphere," as distinct from his verbal characterization and his example, suggests that it is a nonappropriable, and therefore unpaid, factor which gives rise to increasing returns to scale to society though not to the individual producer. At least this is all he needs for the effect he is looking for: a self-policing though nonoptimal competitive situation, where, because the full production functions (i.e., with an axis for rain) are of greater than first degree, the correction of externality via subsidies to promote the creation of favorable atmosphere requires net additions to society's fiscal burden. If this is the crucial consequence of "atmosphere," then it need have no "public" quality. All this would happen even though Smith and Jones were "competing" for the water from the shared water-table under their subsoil, just like bees competing for nectar.

Scitovsky's "two concepts."⁵ Professor Scitovsky, in turn, in his suggestive 1954 article, distinguishes between the statical direct interactions of equilibrium theory and the kinds of pecuniary external economies emphasized in the economic development literature. He classifies the former as consumer-consumer, producer-consumer, and producer-producer interactions, labels the last as external economies and asserts that they are rare and, on the whole, unimportant.

While Scitovsky does not raise the question of what gives rise to such producer-producer interactions, both his examples, and his conclusion that they are of little significance, suggest that he is thinking

4. Formally, Meade denotes "atmosphere" as a situation where the production function, e.g., of farmers takes the form $X_1 = H_1(L_1, C_1)A_1(X_2)$, with L as labor, C as capital and A the atmosphere effect on X_1 of X_2 . The full function exhibits increasing returns to scale but the H function alone, with A constant, is homogeneous of first degree. But why can't this be put in terms of Meade's unpaid factor type function where $X_1 = H_1(L_1, C_1, X_2)$? Example: $X_1 = L_1^a C_1^{1-a} X_2$.

All this has nothing to do with whether $A = A_1 + A_2$ or rather $A = A_1 = A_2$. Unfortunately, the example itself tends to mislead. The fact that exclusion of rain-users (farmers) by producers (timber-growers) is hardly feasible, i.e., that rain is like Type (1), distracts attention from the important point that *if* rain is, as Meade tells us, a public good, then rationing it by price would be inefficient even if it were feasible. (It should be said that Meade concludes his article: "But, in fact, of course, external economies or diseconomies may not fall into either of these precise divisions and may contain features of both of them.")

5. *Op. cit.*

primarily of Type (1): nonappropriability. But this is to ignore public goods — surely a more important cause of interaction. Moreover, by taking full account of these, Scitovsky's "fifth and important case, which, however, does not quite fit into . . . (his) . . . classification . . . , where society provides social services through communal action and makes these available free of charge to all persons and firms," can be made nicely to fall into place.⁶

Samuelson on Types (2) and (3). While the public good model helps to sort out the phenomena Meade lumped under "atmosphere," Samuelson himself emphasizes the analytical bond between indivisibility and public good situations. In both an explicit "summing in" is required of "all direct and indirect utilities and costs in all social decisions."⁷ In Type (2) situations it is the intramarginal consumer's and producer's surpluses associated with various all or nothing decisions "in-the-lump" that have to be properly (interpersonally) weighted and summed, while in Type (3) it is only utilities and costs at the margin that require adding. But, and this is the crucial shared quality of the two categories, both make it necessary to sum utilities over many people.⁸

Exclusion. One more comment may be warranted on the significance, in a public good type situation, of nonappropriability. "Exclusion" is almost never impossible. A recluse can build a wall around his garden, Jones can keep his educated children away from those of

6. *Ibid.*, fn. 3, p. 144. Scitovsky, following Meade, restricts his "first concept" of external economies to phenomena consistent with competitive equilibrium. He treats indivisibilities and increasing returns to scale as belonging to his "second concept" which has to do with disequilibrium, investment decisions, and growth. It is, of course, entirely legitimate to restrict analysis to competitive situations. But the Scitovsky treatment must not be taken to imply that lumpiness is irrelevant to statical analysis of stationary solution points. If one is interested in the statical efficiency of decentralized price calculations, they are crucial. But this is carping. Scitovsky's important contribution lies in emphasizing and clarifying the point first hinted at by P. N. Rosenstein-Rodan that in a world of disequilibrium dynamics pecuniary external economies may play an independent role — one distinct, that is, from simply being an unreliable signal of monopoly troubles (*Economic Journal*, LIII, 1943, 202-11).

7. *Ibid.*, p. 9.

8. There is one qualification to be made: if all public good and increasing returns to scale industries produce only intermediate products, all externalities may cancel out in intra-business-sector transactions. If so, only total revenues and total costs have to be summed. Incidentally, the exposition may misleadingly suggest another symmetry between Types (2) and (3). In a pure Type (3) situation, *if* there are no public producers' goods, then while prices cannot be used to ration the bliss point output-mix, they can be used efficiently to mediate production. In Type (2), on the other hand, *if* all final consumables are divisible, price calculations, while failing in production, will work in exchange. This symmetry breaks down, of course, as soon as one violates, as does the real world, the two "if's."

Smith, etc. But if thereby some people (e.g., the recluse) are made happier and some (e.g., the passers-by) less happy, any decision about whether to "exclude" or not implies an algebraic summing of the somehow-weighted utilities of the people involved. And if the wall requires scarce resources, the final utility sum must be matched against the cost of the wall. When Type (3) blends with indivisibility in production, as it does in the case of the wall, or in the case of a lighthouse, the comparison has to be made between intramarginal totals. Where no lumpiness is involved (e.g., the decibels at which I play my radio) only *MRS* and perhaps *MC* calculations are called for. But the really crucial decision may well be about how much perfectly feasible appropriation and exclusion is desirable.

Arrangements to offset. It is of interest to speculate what, if any, organizational rearrangements could offset the three categories of externality and avoid the need for centrally calculated tax-subsidy schemes.⁹ In concept, Type (1) can be offset by rearrangements of ownership and by "proper" bookkeeping, such as need not violate the structural requirements of decentralized competition. Further, no resort to nonmarket tests would be required.¹

Types (2) and (3) are not so amenable to correction consistent with decentralized institutions. The easiest possible case occurs where increasing returns obtain on the level of single producers'-good plants, much of whose production can be absorbed by a single user firm. Here vertical integration takes care of the problem. Not every process inside a well-run firm is expected to cover its cost in terms of the correct set of internal accounting (shadow) prices. Total profits are the only criterion, and it may pay a firm to build a private bridge between its two installations on opposite sides of a river yet charge a zero accounting price for its use by the various decentralized manufacturing and administrative divisions; the bridge would make accounting losses, yet total company profits will have increased. As long, then, as such integration is consistent with the many-firms requirement for competition, no extra-market tests are required.² The private total conditions: *TR* less *TC*, correctly account for social gain.

9. For illustrative derivation of the formulas for corrective taxes and subsidies in Type (1) situations, see Meade (*op. cit.*).

1. The Emancipation Proclamation could constitute, of course, a substantial barrier.

2. If, however, the "break even" scale of operation of the integrated firm (i.e., where *MC* cuts *AC* from below) is much greater than if the river had not been there to span, or could be spanned by some means of a lower fixed-cost-to-variable-cost ratio, the monopoly problem may simply be "pushed forward" to consumer markets.

Where a producers'-good firm, required to produce at a stage of falling AC , sells to many customer firms and industries, an adding up of all the associated TR 's and TC 's at the precalculated "as if" competitive prices associated with the bliss point would again effectively "mop up" all social costs and benefits.³ But the institutional reorganization required to get correct decentralized calculation involves horizontal and vertical integration, and the monopoly or oligopoly problem looms large indeed. The Type (3) case of a pure *producers'* public good belongs here: only input MRS 's along production functions require summing.

In the general case of a mixed producer-consumer good (or of a pure consumer good) which is "public" or is produced under conditions of increasing returns to scale, it is impossible to avoid comparison of multiperson utility totals. Explicit administrative consideration must be given, if you like, to consumer's and producer's surpluses for which no market-institution tests exist short of that provided by a perfectly discriminating monopolist. But to invoke perfect discrimination is to beg the question. It implies knowledge of all preference functions, while as Samuelson has emphasized,⁴ the crucial game-theoretical quality of the situation is that consumers will not correctly reveal their preferences: it will pay them to "cheat."

Blends. Examination is needed of various blends of Types (2) and (3), such as Sidgwick's lighthouse;⁵ or, for that matter, and as suggested by Samuelson, of blends of public and private goods even where all production functions are fully convex. There are many puzzling cases. Do bridge crossings differ in kind from radio programs? Both involve indivisibility and, where variable cost is zero for the bridge, zero MC 's. The correct price for an extra stroller, as for an extra listener, is clearly zero. Yet bridge crossings have a distinctly private quality: bridges get congested, physical capacity is finite. This is not true of a broadcast. There is no finite limit to the number

3. Assuming that all consumer goods are finely divisible and require no lumpy decisions by consumers.

4. Cf. any of the three "Public Expenditure" articles (*supra*).

5. Sidgwick, by the way, as also Pigou, thought of a lighthouse as of Type (1). It is, of course, "inconvenient" to levy tolls on ships, but it is hardly impossible to "exclude," for instance by means of "serambling" devices (though poor Sidgwick could hardly have known about such things). The point is, rather, that it would be inefficient to do so: the marginal cost to society of an additional ship taking directional guidance from the beacon atop the Statue of Liberty is zero, *ipso* price should be zero. In the case of a lighthouse this is twice true: because the beacon is in the nature of a public good: more for the Queen Mary means no less for the Liberté; and because a lighthouse is virtually an all-fixed-cost, zero variable-cost facility.

of sets that can costlessly tune in.⁶ Radio programs, then, have a public dimension. Yet, in a sense, so do bridges. While your bridge crossing is not my bridge crossing, in fact could limit my crossings, your bridge is my bridge. What is involved here is that most things are multidimensional and more than one dimension may matter.

V. EFFICIENCY, MARKETS AND CHOICE OF INSTITUTIONS

All the above has to do with the statical efficiency of price-directed allocation in more or less idealized market situations. Relevance to choice of institutions depends, of course, on the prevalence of the phenomena which cause externality and on the importance to be attached to statical efficiency. Space precludes extensive discussion of these important issues, but a few casual comments, in the form of *dicta*, are perhaps warranted.

How important are nonappropriability, nonconvexity and public goods? I would be inclined to argue that while nonappropriability is of small import,⁷ the same cannot be said of the other two. True enough, it is difficult to think of many examples of pure public goods. Most things — even battleships, and certainly open air concerts and schools (though not knowledge) — have an “if more for you then less for me” quality. But this is of little comfort. As long as activities have even a trace of publicness, price calculations are inefficient.⁸ And it is surely hard to gainsay that some degree of public quality pervades much of even narrowly “economic” activity.

Lumpiness, in turn, and nonlinearity of the increasing returns sort, while in most instances a matter of degree, and, within limits, of choice, are also in the nature of things. The universe is full of singularities, thresholds and nonproportionalities: speed of light, gravitational constant, the relation of circumference to area, etc. As economists we can cajole or bully engineers into designing processes and installations that save on congealed inputs and give smaller maximal service yields, especially when designing for low-income communities. But the economically perhaps arbitrary, not completely physics-imposed quality of indivisibilities associated with standard designs

6. Richard Eckaus has suggested to me that it is possible to exhaust the space to which the broadcast is limited and that this makes the situation a little more like that of a bridge. Neither of us is entirely satisfied, however.

7. Except for labor skills — and these would take us beyond the bounds of reversible statics.

8. This is not to say that there exist other feasible modes of social calculation and organization which are more efficient.

and ways of doing things should not blind. Nonlinearity and lumpiness are evident facts of nature.⁹

More important, at this level of discourse¹ — though perhaps it hardly need be said — is that statical market efficiency is neither sufficient nor necessary for market institutions to be the “preferred” mode of social organization. Quite apart from institutional considerations, Pareto efficiency as such may not be necessary for bliss.² If, e.g., people are sensitive not only to their own jobs but to other people’s as well, or more generally, if such things as relative status, power, and the like, matter, the injunction to maximize output, to hug the production-possibility frontier, can hardly be assumed “neutral,” and points on the utility frontier may associate with points inside the production frontier.³ Furthermore, there is nothing preordained about welfare functions which are sensitive only to individual consumer’s preferences. As a matter of fact, few people would take such preferences seriously enough to argue against any and all protection of individuals against their own mistakes (though no external effects be involved).

All this is true even when maximization is subject only to technological and resource limitations. Once we admit other side relations, which link input-output variables with “noneconomic” political and organizational values, matters become much more complicated. If markets be ends as well as means, their nonefficiency is hardly sufficient ground for rejection.⁴ On the other hand, efficient markets may

9. Their quantitative significance is, of course, very sensitive to scale, to “size” of markets. This explains the particular emphasis on the role of “social overheads” in low income countries.

1. Where recourse to strategic considerations of feasibility, crucial though they be, is quite out of order.

2. That it is never sufficient is, of course, well known. Of the infinite Pareto-efficient configurations at best only one: that which gives the “right” distribution of income in terms of the *W*-function that is to count, has normative, prescriptive significance. Moreover, most interesting *W*-functions are likely to be sensitive to “noneconomic” factors, such as are, if not inconsistent, at least extraneous to Paretian considerations. Where such additional values of a political or social nature are separable from input-output values (i.e., where the two sets can be varied independently of each other) one “can” of course separate the overall *W*-function into a “political” and an “economic” component and maximize separately over each.

3. This is different from the usual case of consumer sensitivity to the input-output configuration of producers, e.g., factory soot or a functional but ugly plant spoiling the view. Such joint-product “bads” can be treated as inputs and treated in the usual Paretian fashion. It is a different matter that their public quality will violate duality, hence render market calculation inefficient.

4. This is too crude a formulation. It is not necessary that markets as such be an “ultimate” value. Political and social (non-output) values relating to the configuration of power, initiative, opportunity, etc., may be so much better served

not do, even though Pareto-efficiency is necessary for bliss. Even with utopian lump-sum redistribution, efficiency of the "invisible hand" does not preclude preference for other efficient modes of organization, if there be any.⁵

Yet when all is said, and despite the host of crucial feasibility considerations which render choice in the real world inevitably a problem in the strategy of "second best," it is surely interesting and useful to explore the implications of Paretian efficiency. Indeed, much remains to be done. There is need, in particular, for more systematic exploration of the inadequacies of market calculation in a setting of growth.⁶

by some form of nonefficient market institutions than by possible alternative modes of more efficient organization as to warrant choice of the former. The analytical point, in all this, is that the outcome of a maximization process and the significance of "efficiency" are as sensitive to the choice of side-conditions as to the welfare-function and that these need be "given" to the economist in the same sense that a welfare function has to be given.

5. The above is still strictly statical. For related dynamical problems, e.g., possible conflict between one-period and intertemporal efficiency, cf., "On Capital Productivity, Input Allocation and Growth," this *Journal*, LXXI (Feb. 1957).

6. The development literature on market failure, while full of suggestive insight, is in a state of considerable confusion. Much work is needed to exhaust and elucidate the seminal ideas of Young, Rosenstein-Rodan, Nurkse and others. For important beginnings, see Scitovsky (*op. cit.*), M. Fleming, "External Economies and the Doctrine of Balanced Growth," *Economic Journal*, LXV (June 1955), and Fellner (*op. cit.*).

The view that we should not turn social historian or what not, that the logic of economizing has some prescriptive significance, rests on the belief that narrowly "economic" efficiency is important in terms of many politically relevant *W*-functions, and consistent with a wide variety of power and status configurations and modes of social organization. On the other hand, some may feel that the very language of Paretian welfare economics: "welfare function," "utility-frontier," in relation to choice of social institutions, is grotesque. What is at stake, of course, is not the esthetics of language, on which I yield without demur, but abstraction and rigorous theorizing.

A CRITIQUE OF CONCEPTS OF WORKABLE COMPETITION

By STEPHEN H. SOSNICK*

Introduction, 380. — I. The nature of the theory of workable competition, 383. — II. The several concepts of workability, 386. — III. Performance norms and value judgments, 391. — IV. Necessary vs. sufficient conditions, 395. — V. Remediability vs. categorical norms, 402. — VI. The workability criterion reformulated, 410. — Appendix, 415.

This paper reviews the principal questions which have arisen as economists have developed the theory of workable competition. The answers that are suggested may be summarized as follows:

The theory of workable competition is best understood as an attempt to indicate what practically attainable states of affairs are socially desirable in individual capitalistic markets. In its development, the 1930's are a sort of benchmark. Since then, two points have received more general emphasis. One is that perfect competition is not a reliable basis for normative appraisal of actual markets. The other is that it is necessary to formulate explicitly a criterion of workable competition; that is, a criterion whose greater fulfilment implies a socially better situation, and whose complete fulfilment is necessary and sufficient for achievement of what the public interest demands.

All useful criteria of workability — including those specifying adequacy of alternatives and absence of market power — amount to a set of mutually consistent norms for certain individual market characteristics. The various characteristics of a market have different normative significance. This has usefully been emphasized by grouping them into three somewhat overlapping categories — performance, conduct, and structure.

Eighteen writers have proposed sophisticated criteria of workability. These can be summarized in a rather long list of norms for performance, conduct, and structure. Fundamentally the criteria

* University of California, Davis. The author is indebted to Irma G. Adelman, Joc S. Bain, and William J. Baumol for their helpful comments and to Joe S. Bain and Andreas G. Papandreou for guiding the dissertation from which the present paper is largely derived. The dissertation, "Contemporary Norms for Market Structure and Behavior: A Critical Appraisal," was deposited in the library of the University of California, Berkeley, in 1956. The present article is research paper No. 169 of the Giannini Foundation.

proposed differ in two respects. First, they demand a satisfactory situation in different combinations of the three categories — just performance, or both conduct and structure, etc. — and, within a mutually chosen category, in different combinations of its component elements or dimensions. Second, they demand different achievements for given individual dimensions to be satisfactory.

Performance norms summarize the relevant “economic” and “political” values. These cannot be bypassed by including only structure or conduct norms. The values can be handled somewhat objectively by viewing performance attainments as instrumental to certain generally acceptable social goals, and the importance of structure and conduct conditions as deriving from their net implications for performance. A presumption in favor of buyers’ sovereignty helps to resolve what is inherent — interaim and interpersonal comparisons.

Performance is what is of ultimate importance here. Whether it is satisfactory cannot be inferred from fulfilment of requirements for structure and conduct; satisfactory performance requires an exercise of managerial discretion and a sense of social responsibility. Consequently, structure-conduct accomplishments should not be sufficient for workability. But they should be necessary. Performance often can be appraised only roughly and manipulated only indirectly, so that it is not clear whether past and prospective performance could be improved or assured unless recent and potential revisions in structure and conduct are considered. Price-fixing should not be tolerated merely because performance is stressed and already seems favorable, or because collusion’s effects in a particular market cannot be traced with certainty, or because it has *some* favorable implications for performance. Satisfactory performance, conduct, *and* structure should be necessary and sufficient for workability.

Traditionally the “monopoly problem” encompasses only certain aspects of performance, conduct, and structure. However, these are interdependent with other market characteristics that represent and influence the achievement of satisfactory performance. All must be appraised simultaneously. The “monopoly problem” makes sense not as the evils of high concentration, but as the whole problem of achieving a socially optimum market situation. The workability criterion should demand a satisfactory state of *all* malleable dimensions of performance, conduct, and structure.

The criterion used to delineate the area of appraisal is not critical. The more convenient of demand cross-elasticities or technological considerations may be used. With either, the market’s condition

can be appraised only if analysis is extended to connections with outside products and producers.

Certain well-known structure and conduct attributes almost invariably tend to worsen performance. These the workability criterion should condemn categorically. But usually the feasibility and desirability of specific structure and conduct characteristics, like low concentration, depend on the degree to which many others are or can be present, and on a variety of considerations which vary from market to market. Likewise, the same performance results cannot and should not be realized in different kinds of market situations. Guesswork and subjectivity cannot be avoided in appraising these characteristics. But gross errors can be, by appraising them with the help of a body of contingent predictions and qualified judgments, and in the light of performance effects, needs, and compromises in a particular market. The norms for these characteristics should demand merely that they be in a state which, in view of economic and legal necessities, has no feasible and preferable alternative. No qualification is warranted by the fact that cause or blame for a deficiency may not be attributable to the enterprises concerned.

These norms can be aggregated to provide a goal for market situations, a standard of "adequate" conditions, and a basis for rating actual cases. A market should be adjudged "almost workable," "workable" or "optimum" according to the frequency and extent that its individual performance, conduct, and structure attributes depart from being as favorable as unavoidable circumstances permit. This criterion differs from previous suggestions in emphasizing performance, conduct *and* structure, in including *all* malleable dimensions of each category, in demanding specific attainments for a few dimensions combined with unimprovability for those remaining, and in appraising improvability by reference to *all* significant effects of *all* practicable remedies.

The "optimum" rating requires attainment of specific conditions only for the few market characteristics which are desirable generally. This permits the concept to apply to all sorts of "unregulated" capitalistic markets, but without presupposing optima elsewhere. Underlying a rating are a variety of subjective judgments of value and fact. The part of these pertaining to social goals makes the concept useful to public policy. However, "noneconomic" considerations make it undesirable — indeed, impossible — for the law to enforce workable competition and solve the problem of monopoly.

A catalog of market dimensions and norms is suggested in the appendix.

I. THE NATURE OF THE THEORY OF WORKABLE COMPETITION

The theory is probably best understood as an attempt to formulate certain normative standards. The norms are to indicate what practically attainable states of affairs are socially desirable in the condition of individual capitalistic markets.¹ Fulfilment of the norms is to represent a satisfactory solution to the problem of monopoly.

It would be misleading to suggest that the theory dates from 1933, when many of its fundamental notions were presented by Chamberlin, or from 1940, when it was elaborated and christened by J. M. Clark.² Contemporary normative analysis of the monopoly problem does not sharply depart from the views of earlier writers.³ However, matters of degree are important, and the recent more general emphasis on two fundamental points makes it helpful to think of the 1930's as something of a benchmark.

A. *Perfect Competition Rejected as a Norm*

The first point is quite generally accepted,⁴ although — as will appear later — its radical implications are not yet fully digested. The point is that the set of market structure and conduct attributes which define "perfect competition" constitute individually and collectively neither a normative ideal nor a satisfactory basis for appraising actual market conditions.⁵

What came to be called "perfect competition" involves an effectively infinite number of relatively small buyers and sellers of a standardized product, each rationally seeking his maximum advantage, independently of other traders and free of public "controls," in a continuous market with "complete" and costless knowledge, access, and mobility. The arguments against idealizing atomistic competition, perfect and imperfect, cover a variety of issues — nonoptimum

1. Compare E. S. Mason, "Workable Competition Versus Workable Monopoly," chap. 18 of Mason's *Economic Concentration and the Monopoly Problem* (Cambridge, 1957), p. 382.

2. E. H. Chamberlin, *Theory of Monopolistic Competition*. J. M. Clark, "Toward a Concept of Workable Competition," *American Economic Review*, XXX (June 1940), 241-56.

3. See S. Peterson, "Antitrust and the Classic Model," *American Economic Review*, XLVII (March 1957), 60-78.

4. See, e.g., the *Report of the Attorney General's National Committee to Study the Antitrust Laws* (Washington, 1955), p. 316.

5. Perfect competition remains a very useful concept for certain analytical comparisons, as in defining "entry barriers"; for a first approximation to certain descriptive results, as in deciding what profits tend to be with low concentration; for aggregative analyses abstracting from unpredictable market irregularities, as in appraising the Pigou effect; and for an indication of what structure-conduct elements may have especial influence and what performance results especial pertinence in the operation of actual markets.

conditions elsewhere, economies of scale, poor management, capital shortage, requisites and effects of exploration and of research and development, external economies and diseconomies, dispersion and individuality in traders and variety in outputs, immobile excess capacity, chronic distress, unfair tactics, depression-spreading, incorrect and inconsistent expectations, atoms blundering about a moving equilibrium.⁶ In addition, it is recognized that the extremes which define atomistic and otherwise-perfect competition tell us nothing about desirable gradations in even the few dimensions to which they refer; yet market characteristics are scaled, not either-or, so that the choices open to normative appraisal relate not to whether, but to how much there should be of concentration, bigness, heterogeneity, price flexibility, etc.⁷

These arguments are well known and need not be detailed here. What they amount to is recognition that the perfectly competitive structure and conduct are unattainable in any real market, that closer approximations to them may entail worse performance than more distant, and that the closest possible approximation would entail actual and even equilibrium performance of dubious desirability.⁸

This conclusion is sometimes obscured by an unfortunate use of language. The label "monopolistic" is often attached to markets and market characteristics whose concentration, heterogeneity, and perhaps other attributes depart from perfect competition. In addition, "monopolistic" in ordinary and legal usage describes something which is socially undesirable, while "competitive" carries overtones of approval. The result is that "competitive" and "monopolistic"

6. The development of these arguments in the 1930's can be sketched as follows: J. M. Clark, *Studies in the Economics of Overhead Costs*, e.g., p. 460; Chamberlin, *Theory of Monopolistic Competition*, *op. cit.*, the especially pertinent arguments of which Chamberlin has recently restated in "Product Heterogeneity and Public Policy," *American Economic Review*, XL (May 1950), 85-92, and in his paper in the collection *Monopoly and Competition and Their Regulation* (London, 1954), which he also edited; R. F. Kahn, "Some Notes on Ideal Output," this *Journal*, XLV (Mar. 1935), 29; D. H. Wallace, "Monopolistic Competition and Public Policy," *American Economic Review*, XXVI (Mar. 1936), 80; J. M. Clark, *Social Control of Business*, 2d ed., pp. 135-36; Clark, "Toward a Concept of Workable Competition," *op. cit.*, pp. 241-56; J. A. Schumpeter, *Capitalism, Socialism, and Democracy*, 3d ed., e.g., p. 106; T. O. Yntema, "Competition as a Norm of Economic Behavior," *Journal of Business*, XIV (July 1941), 271-76.

7. See, e.g., M. A. Adelman, "Integration and Antitrust Policy," *Harvard Law Review*, LXIII (1949), 48.

8. The following provides an interesting prebenchmark contrast: "... imperfect though actual competition is, one cannot deny that it *tends* to bring about the results which perfect competition would bring about. It is a force operating in the same direction, although with diminished power and velocity." M. W. Watkins, *Industrial Combinations and Public Policy*, pp. 103-4.

can be used only at the risk of evoking a descriptive connotation of closeness to perfect competition *joined with* an emotive connotation of approval or disapproval. The point here, however, is that such a distinction between competition and monopoly is not helpful in appraising the state of affairs in actual markets. Markets are invariably monopolistic in this sense, and without other information one does not know whether making them more or less monopolistic would tend to improve their performance. If a reduction (or an increase) in a market's concentration proves to be desirable, it is better to say so directly, not to suggest that the conclusion follows from an ideal of perfect competition by saying — as is common — that the market is “not sufficiently competitive.”⁹

B. Explicit Standards Required

The second matter of emphasis which may be said to characterize the theory of workable competition is an effort to be quite explicit in setting forth the normative standards whose fulfilment is to represent a satisfactory solution to the monopoly problem.

It is not enough to suggest ground for confidence “that competition usually would . . . be extremely effective among as few as five or six competitors.”¹ What “effective competition” means must be formulated clearly; and standards must be set forth which provide guidance and assessment for both “the antitrust laws, which are enforced by a series of cases against designated enterprises, not by general evaluations,” and also for other “types of public policy in which big business generally [and other market conditions too] would be encouraged or curbed.”² It is thus not enough even to formulate certain necessary conditions of workable competition — like no pricing collaboration — although these may be all that the law can impose. The task is to set forth the practically attainable conditions whose presence is both necessary and sufficient in a particular segment of the economy for it to be doing or being what the public interest may reasonably demand of it, and whose greater fulfilment would imply greater social benefit. This is the definition of “workable competition” or, to abbreviate, the workability criterion.

9. Indeed, the conclusion would not follow even if atomism were desirable and this implied that looser oligopoly was better than tighter. As Chamberlin notes, “. . . to break up . . . [large] units would . . . almost certainly involve an increase in product differentiation . . . [and perhaps] in the aggregate of monopoly power. . . .” *Theory of Monopolistic Competition*, 7th ed. (Cambridge, 1956), p. 214, n. 1.

1. J. B. Clark, *Essentials of Economic Theory*, p. 201; quoted by Peterson, *op. cit.*, p. 71.

2. C. D. Edwards, *Big Business and the Policy of Competition* (Cleveland; Press of Western Reserve University, 1956), pp. 20–21, 26.

Seeking explicit standards justified without idealizing perfect competition, we turn now to a review of the criteria of workability which have recently been suggested.

II. THE SEVERAL CONCEPTS OF WORKABILITY

In a field so ill bounded and well trod, it is difficult to decide which writers' proposals should be included. The most I can say is that "reasonably sophisticated" tests of a market's workability have been proposed to my knowledge by eighteen writers, who will be named shortly. Among the eighteen, views have changed with time, and analysis has differed in many and subtle ways. The many variations could be identified in a paper this size only at the sacrifice of critical appraisal. I will emphasize only the two points of difference which seem fundamental.³ The first, occupying Sections II, III, and IV, concerns which of a market's characteristics should determine its workability. The second, occupying Sections V and VI, concerns what to demand of the characteristics that are relevant.

A. *Structure, Conduct, and Performance Distinguished*

Obviously, not all market characteristics have equal significance. This is recognized by grouping the various dimensions of a market into several categories — three most commonly. Unfortunately, headings are seldom defined and usage is not uniform even for individual writers.⁴ "Structure" will be used here to refer to characteristics which constitute a market's patterns, status, composition. "Conduct" will refer to characteristics which are enterprises' actions,

3. A more detailed exposition will be found in my dissertation, *op. cit.*, chap. 3.

4. See, e.g., J. S. Bain, "The Normative Problem in Industrial Regulation," *American Economic Review*, XXXIII (Mar. 1943), 55; "Price and Production Policies," chap. 4 in H. S. Ellis (ed.), *Survey of Contemporary Economics*, I, 142-43, 160-61; "Workable Competition in Oligopoly," *American Economic Review*, XL (May 1950), 38, 44-45. C. D. Edwards, *Maintaining Competition*, pp. 9-10; "Public Policy and Business Size," *Journal of Business*, XXIV (Oct. 1951), 285-86. C. E. Griffin, *An Economic Approach to Antitrust Problems* (No. 441 in the American Enterprise Association's series, "National Economic Problems"; New York, 1951), p. 59. E. S. Mason, "Competition, Price Policy, and High-Level Stability," in *Economic Institute on Pricing Problems and the Stabilization of Prosperity* (U.S. Chamber of Commerce, Washington, 1947), pp. 21-22; "Current Status of the Monopoly Problem in the United States," *Harvard Law Review*, LXII (June 1949), 1267-82; "Economists View Bigness in Business," *Commercial and Financial Chronicle*, CLXXI (Jan. 26, 1950), 25; "Methods of Developing a Proper Control of Big Business," *Proceedings of the Academy of Political Science*, XVIII (Jan. 1939), 44. "Price and Production Policies of Large-Scale Enterprise," *American Economic Review*, XXIX (Mar. 1939), 66, 73-74. D. H. Wallace, "Industrial Markets and Public Policy: Some Major Problems," *Public Policy*, I, 115, 128.

dealings, or tactics. "Performance" will refer to dimensions which represent the realization of normatively significant "economic" results.⁵ Conduct and performance together are sometimes called "behavior."

It is important to notice that "structure," "conduct," and "performance" have been defined in a way which allows individual market characteristics to fall under more than one heading. Structure and conduct overlap because certain events can usefully be viewed as patterns in some contexts and as processes in others. Thus, product variety can be viewed either as the presence of heterogeneity in outputs or as producers' active creation of differently-received outputs — product differentiation. Performance and structure-conduct overlap because certain patterns or practices not only influence performance, but simultaneously bear directly on what the observer sees as aims for a good economy, and so have normative significance — performance standing — in themselves. Thus, product variety can be interpreted also as an allocation of resources which adjusts to buyers' diverse budgets, wants, and locations.

It is also important to notice that this separation into structure, conduct, and performance corresponds only imperfectly to the direction of causation. Causation will operate in various directions within and among the several layers. A structure which includes, say, high concentration of supply may promote certain conduct, say price stabilization; but conduct, say combination, may alter structure. Structure and conduct together may generate certain performance, say excessive profits, but the profits may in turn attract entry which modifies concentration or pricing practices.

Basically, however, the categories correspond to steps in a means-ends argument. They thereby form the basis of different analytical approaches to the concept of workable competition. Indeed, it can be asserted that all useful approaches to the concept ultimately reduce to standards for a market's structure, conduct, or performance.

B. Two Specious Tests

The point applies to the legal shibboleths — no power to raise prices or exclude competitors,⁶ on which, wisely, none of our eighteen writers rely. Taken literally, these criteria demand perfect competition, and condemn it even if demand and price rise. If not taken literally, the criteria evaporate into a search for standards which can

5. The qualification "economic" is needed here, despite its vagueness, unless we are prepared to cover everything involved to which "the community" assigns value — the presence of religious symbols, for example.

6. See, e.g., *Report of the Attorney General's Committee*, *op. cit.*, p. 44.

distinguish and assess various types and degrees of power according to their origin, strength, duration, or exercise.⁷ The search can turn nowhere except to norms for aspects of a market's structure, conduct, and performance — aspects like output heterogeneity, profit levels, proximity of substitutes, and predation and coercion.⁸

Another specious test — this one suggested by several of the writers, Adelman, Edwards, and Wilcox — proposes in effect that competition be considered workable if there are available to traders an adequate number of genuine alternatives.⁹ In itself, this criterion merely rephrases the original question. Competition is just the process of providing options or alternatives.¹ Alternatives always exist, since competition "merges into substitution,"² and "one could always use something else in place of aluminum."³ What is significant for the workability of any one market might be the number, kind, or importance of the price or product options or their providers or receivers. The criterion is not operational until the meaning of "alternatives" and the tests of their adequacy are made concrete. When this is done, it is evident that the notion of adequate alternatives is merely a different label for a workability criterion which in fact is composed of structure, conduct, or performance norms.⁴

7. *Ibid.*, pp. 48–55.

8. Compare Adelman, "Integration and Antitrust Policy," *op. cit.*, p. 48; Mason, *Economic Concentration*, *op. cit.*, pp. 387–88.

9. Adelman declares that for workable competition, "A sufficient number of alternatives open to any given buyer or seller are necessary. . . ." M. A. Adelman, "Business Size and Public Policy," *Journal of Business*, XXIV (Oct. 1951), 272; "Effective Competition and the Anti-Trust Laws," *Harvard Law Review*, LXI (Sept. 1948), 1298, 1303–4. Edwards asserts that competition, "As a safeguard against extremes of exploitation, lethargy, and incompetence" "consists in access by buyers and sellers to a substantial number of alternatives and in their ability to reject those which are relatively unsatisfactory." C. D. Edwards, *Maintaining Competition*, *op. cit.*, pp. 9, 124. Wilcox states that competition is workable if there are available to buyers and sellers, "Genuine alternatives in policy" among their respective sources of supply and demand sufficient to enable them, by shifting their dealing from one to another, "Substantially to influence quality, service, and price." C. Wilcox, *Competition and Monopoly in American Industry* (Monograph 21, U.S. Temporary National Economic Committee, Washington, 1940), pp. 8–9, 19–20; "A Program to Promote Competition," chap. 16 of G. W. Stocking and M. W. Watkins, *Monopoly and Free Enterprise*, p. 540.

1. Study the definitions of J. M. Clark, *Alternative to Serfdom*, p. 70, and C. E. Griffin, "Economic Objectives and Antitrust Policy," in *Lectures on Federal Antitrust Laws* (Ann Arbor, Michigan; 1953), pp. 31–33.

2. Clark, *Alternative to Serfdom*, *op. cit.*, p. 70.

3. G. J. Stigler, "The Extent and Bases of Monopoly," *American Economic Review*, XXXII (June 1942), 3, fn. 5.

4. Adelman refers to alternatives which include the types of goods — for example, stripped versus begadgetted models — and appears to judge their adequacy by conduct criteria. "Effective Competition," *op. cit.*, p. 1303. Edwards refers to actual and potential alternative customers and suppliers and to alterna-

The question of what conditions are necessary and sufficient for workability, then, becomes the question of which of the categories, and which of their component dimensions, should be appraised by the workability criterion. The criterion itself will be an aggregate of what may be called "dimension norms" — normative standards for individual aspects of structure, conduct, or performance. Debate has centered on which norms to include.

C. Proposed Dimension Norms

Certain structure dimension norms are suggested by Baum, Clark, Edwards, Lewis, Markham, Mason, Mund, Smith, Stigler, Stocking and Watkins, and Wilcox: (1) A large or an appreciable number of traders, or several at least, none dominant; or as many or at least as many as scale economics permit.^{5c,d,e,g,h,i,k,n,o,q} (2) Moderate

tive policies on price, output, kinds of goods, and quality; he lists a number of structure-conduct requisites of adequacy. "Can the Antitrust Laws Preserve Competition?" *American Economic Review*, XXX (Mar. 1940), 11-12; *Maintaining Competition*, *op. cit.*, pp. 9, 124. Wilcox refers to genuine alternatives in policy; for their adequacy he enumerates a series of structure and behavior tests which are far more stringent than his other suggestion, since almost anywhere traders would, if many shifted, substantially influence quality and price. *Competition and Monopoly* . . . , *op. cit.*, pp. 8-9.

Others of the eighteen writers have used the expression. Baum refers to the number of real alternatives of price and quality; he would judge them by structure and conduct tests. W. C. Baum, *Workable Competition in the Tobacco Industry* (Ph.D. dissertation, Harvard, 1949), pp. 16-18, 61-62. Griffin speaks of "Alternative sources of goods and services," but suggests "Chief reliance upon performance." C. E. Griffin, *An Economic Approach to Antitrust Problems* *op. cit.*, p. 60; *Enterprise in a Free Society*, p. 300. Mason speaks of the number of alternative buyers and sellers independently probing the possibilities in the situation, and argues that performance must be the test of adequacy. E. S. Mason, "Current Status of the Monopoly Problem," *op. cit.*, pp. 1280-82. Smith speaks of existing and potential alternatives among buyers and sellers and among goods, services, and markets; he suggests what are essentially criteria for conduct that *tends to produce* better alternative products. B. Smith, "Effective Competition: Hypothesis for Modernizing the Antitrust Laws," *New York University Law Review*, XXVI (July 1951), 412-16, 429, 436; compare the often identical language in *Effective Competition* (Report to the Secretary of Commerce by his Business Advisory Council, Dec. 18, 1952). An exception to the transformation of an alternatives approach into one using performance, conduct, or structure norms seems to arise from Baum's suggestion that "the testimony of consumers of the product offers a significant, although not rigorous, test" of the adequacy of alternatives. *Workable Competition in Tobacco*, *op. cit.*, p. 18. Buyers' testimony may indeed help to *indicate* whether items like product variety are satisfactory. But it does not itself *replace* those evaluative factors, unless buyers' contentment is a value over and above the objective conditions which ordinarily produce it and other things too. If so, merely another performance dimension has been introduced.

5. a. M. A. Adelman, "The A & P Case: A Study in Applied Economic Theory," this *Journal*, LXIII (May 1949), 247-53; "Business Size and Public Policy," *op. cit.*, p. 274; "Effective Competition," *op. cit.*, pp. 1303-4; "Integra-

and price-sensitive quality differentials.^{c,d,e,k,n,o,q} (3) No artificial handicaps on mobility.^{c,d,e,k,m,n,q} (4) Adequate access to information.^{c,e,k,n,q} (5) Some uncertainty whether a reduction in price will

tion and Antitrust Policy," *op. cit.*, p. 47; "The Large Firm and Its Suppliers," *Review of Economics and Statistics*, XXXI (May 1949), 114-17.

b. J. S. Bain, *Barriers to New Competition* (Cambridge, 1956), pp. 210, 218; *The Economics of the Pacific Coast Petroleum Industry*, I, 2-3; II, 1, 240; III, 122; "The Normative Problem," *op. cit.*, pp. 55-56; "Price and Production Policies," *op. cit.*, p. 169; "Relation of the Profit Rate to Industry Concentration," this *Journal*, LXV (Aug. 1951), 294, fn. 4; "Workable Competition in Oligopoly," *op. cit.*, pp. 36-37.

c. W. C. Baum, *Workable Competition in Tobacco*, *op. cit.*, pp. 16-19, 61-62.

d. J. M. Clark, *Alternative to Serfdom*, *op. cit.*, pp. 70-72; *Guideposts in Time of Change*, pp. 115, 143-45; "The Orientation of Antitrust Policy," *American Economic Review*, XL (May 1950), 95, 104; *Social Control of Business*, *op. cit.*, p. 125. "Toward A Concept of Workable Competition," *op. cit.*, pp. 243-56.

e. C. D. Edwards, *Maintaining Competition*, *op. cit.*, pp. 10, 128-29.

f. C. E. Griffin, *Economic Approach to Antitrust Problems*, *op. cit.*, pp. xiii, 62-63; "Economic Objectives and Antitrust Policy," *op. cit.*, pp. 38-40.

g. B. W. Lewis, untitled, *American Economic Review*, XXXIX (June 1949), 705-8.

h. J. W. Markham, *Competition in the Rayon Industry*, pp. 204-7; "Merger Policy Under the New Section 7," *Virginia Law Review*, XLIII (May 1957), 492.

i. E. S. Mason, "The Antitrust Laws: A Symposium," *American Economic Review*, XXXIX (June 1949), 713; "Competition, Price Policy, and High-Level Stability," *op. cit.*, pp. 29-31; "Current Status of the Monopoly Problem," *op. cit.*, pp. 1280-82; "Economists View Bigness," *op. cit.*, p. 25; "The New Competition," *Yale Review*, Aug. 1953, pp. 41-42.

j. R. S. Meriam, "Bigness and the Economic Analysis of Competition," *Harvard Business Review*, XXVIII (Mar. 1950), 125-26; "The Sherman Antitrust Act and Business Economics," in *A New Look in Antitrust Enforcement Trends*, pp. 97-98; a letter to the present writer, dated Sept. 30, 1954.

k. V. A. Mund, *Government and Business*, pp. 71, 73, 174a; "Monopolistic Competition Theory and Public Price Policy," *American Economic Review*, XXXII (Dec. 1942), 728-31, 741-42.

l. A. R. Oxenfeldt, *Industrial Pricing and Market Practices*, pp. 91-92, 557-59.

m. B. Smith, "Effective Competition," *op. cit.*, pp. 412-19, 429, 436-37, 441.

n. G. J. Stigler, "Extent and Bases of Monopoly," *op. cit.*, pp. 2-4; "Mergers and Preventive Antitrust Policy," *University of Pennsylvania Law Review*, CIV (Nov. 1955), 177-83; *Testimony* before the Subcommittee on the Study of Monopoly Power of the Committee on the Judiciary, House of Representatives, 81st Congress, 2d Session, Serial No. 14, Part 4A, pp. 120, 996.

o. G. W. Stocking and M. W. Watkins, *Monopoly and Free Enterprise*, *op. cit.*, pp. 97, 101, 102, 108, 506. G. W. Stocking, "The Rule of Reason, Workable Competition, and Monopoly," *Yale Law Journal*, LXIV (July 1955), 1161; G. W. Stocking, "The Attorney General's Committee's Report: The Businessman's Guide Through Antitrust," *Georgetown Law Journal*, XLIV (Nov. 1955), 6, 11.

p. D. H. Wallace, "Industrial Markets and Public Policy: Some Major Problems," *op. cit.*, pp. 78, 99-100, 110, 126-27.

q. C. Wilcox, *Competition and Monopoly*, *op. cit.*, pp. 8-9, 19-20, 307; "A Program to Promote Competition," *op. cit.*, 540, 543, 552, 554, 555, 563.

be met.^d (6) Absence of legal restrictions.^{e,i,k,m,q} (7) Continual opening of fresh areas and types of competitive contact.^d

Several conduct norms are suggested by Adelman, Baum, Clark, Edwards, Mason, Meriam, Mund, Oxenfeldt, Smith, Stigler, Stocking and Watkins, and Wilcox: (1) Firms should strive in rivalry, pursuing their independent judgment and responding without collusion to considerations of profit and loss.^{a,c,d,e,i,j,k,l,m,n,o,q} (2) Firms should not shield permanently inefficient rivals, suppliers, or customers.^j (3) There should be no unfair, exclusionary, predatory, or coercive tactics.^{e,j,k,m,o,q} (4) Some forms of discrimination should or should not occur.^{a,d,k} (5) Sales promotion should not be misleading.^{a,c,d} (6) Buyers should react fairly rapidly to differential offerings.^d

Various performance norms are suggested by Bain, Baum, Clark, Griffin, Lewis, Markham, Mason, Meriam, Oxenfeldt, Smith, Stocking and Watkins, Wallace, and Wilcox: (1) Operations should be efficient.^{b,c,f,h,i,l,m,o,p,q} (2) Promotion expenses should not be excessive.^{b,i,l} (3) Profits should be at levels which reward investment and efficiency and induce innovation.^{b,d,f,i,j,o,p,q} (4) Output should be consistent with a good allocation of resources.^{b,m,o,p,q} (5) Prices should not intensify the cyclical problem.^{b,d,m,p,q} (6) Quality should conform to consumers' interests.^{l,m,q} (7) Opportunities for better products and techniques should not be neglected.^{b,c,f,g,h,i,j,l,m,o,p,q} (8) Conservation should not be disregarded.^{b,c,m} (9) Success should accrue to sellers who give buyers more of what they want.ⁱ (10) Entry should be as free as the nature of the industry permits.^f (11) The industry should aid in the national defense.^{f,j} (12) Excessive political and economic power should not rest in the hands of small groups.^{c,i,l,m,q} (13) Employees' welfare should not be neglected.^{l,m}

From the citations here it can be seen that many of the writers include no dimension norms of one or even two of the performance, conduct, and structure categories, and that all of the writers omit various norms of even the categories they emphasize. Partly the omissions can be explained by the vagueness of the distinctions among both the three categories and their individual dimensions. But principally the explanation for this first fundamental point of difference lies in real differences of ideology, theory, and opinion. Our concern now is to reconsider which dimensions' condition should determine workability. To do so, we must inquire into the role that value judgments will play.

III. PERFORMANCE NORMS AND VALUE JUDGMENTS

The workability criterion, we shall see, would not necessarily neglect certain "political" values if it were confined to performance

tests, nor could it avoid certain "economic" values by being confined to norms for structure and conduct. These values should be explicit, and can be handled somewhat objectively by viewing them as instrumental to certain generally acceptable social goals.

A. Performance Summarizes Values, Which Are Unavoidable

Performance is, by definition, what is of ultimate "economic" importance in and from markets. If the workability criterion consisted of norms for all performance dimensions, but only them, it would not be objectionable on the ground that it forgot that the presence of more than one source of supply or demand, or the absence of organized barriers to entry, can be "conceived as the diffusion of economic power," "valued for its own sake."⁶ An increase in concentration or entry barriers does not always imply an increase of unwarranted sorts of influence over other persons' actions, since among other things, certain power positions offset others and entry may occur by diversification. But even if so, the issue of concentration of economic power would be covered with other values as dimensions of performance.⁷

Conversely, while questions of values would be summarized in performance norms, the questions would not be avoided — as some might like — by confining the workability criterion to nonperformance aspects of structure or conduct. A rational decision that some condition of structure or conduct was desirable would require the judgment that its various performance implications in the net were favorable. For example, whether it is desirable to have a number of sources of supply depends on whether and to what extent its incremental "political" advantages would be bought at the price of the market's becoming unstable, destructive, inefficient, or stagnant. If we refuse to inquire explicitly into the compromises that are appropriate in the several performance dimensions, we must necessarily presume answers to the underlying questions which we refuse to ask.

The only alternative is to postpone the normative decisions by presenting purely descriptive conclusions. These would be in the form, *A* — say atomistic concentration — in structure-conduct tends to produce the composite *B* — say frequent failures, zero profits, optimum scale, etc. — in presumptively identified performance. This we could not settle for without abandoning the effort to develop a theory of workable competition. Whether the workability criterion

6. Edwards, *Big Business*, *op. cit.*, pp. 2-4.

7. Contrast Edwards, *Big Business*, *op. cit.*, p. 115 (but see p. 105); Mason, "Current Status of the Monopoly Problem," *op. cit.*, pp. 1266-67.

sets forth structure, conduct, or performance norms, performance standards are needed, deriving from a generally acceptable set of social goals.

B. Three Obstacles to Deriving Performance Norms

Derivation of performance norms — for example, the oversimplified norm that profits should equal zero — consists in establishing their connection with more basic aims — say aims for greater per capita production, an advantageous composition of output, an equitable distribution of economic freedom, a favorable influence on the achievement of “noneconomic” values. It must be shown that the more basic aims would be furthered if the norms were accepted as principles for appraising or choosing among alternatives in market performance, and consequently as principles which determine the content of standards that choose among alternatives in conduct and structure. The more basic, focal aims are simultaneously ends relative to the performance which affects them, and means relative to whatever goals in turn justify them. The regression of ethical premises continues until some goals — perhaps freedom and happiness — are accepted as principles of choice which themselves require no justification — accepted with cognizance that goals may lose their abstract attractiveness when it is realized what means must be employed and what other ends sacrificed to advance them. Focal aims, and performance norms, cannot be independent values. They oppose, limit, modify, give meaning to each other. They must each be so qualified as to call for the greatest net contribution to the social goals which together they are supposed to promote.⁸

Let us assume that a generally acceptable and sufficiently meaningful set of focal aims can be identified. Even so, three obstacles to the derivation of performance norms come immediately to mind, and several more will appear later. There is, first, considerable room for speculation and dispute on the factual question of what effects in the direction of particular aims — say progress or equity — would be produced by alternative states of particular dimensions of performance — for example, by “high” or “low” profits. Secondly, a need inevitably arises to choose among alternatives — say in advertising controls — which further some aims as against others — perhaps aims emphasizing freedom as consumers instead of freedom as producers or citizens.

8. Available discussions of social goals are too numerous to cite here. Two of the most valuable are H. M. Oliver, *A Critique of Socioeconomic Goals* (Bloomington, Indiana; 1954), and J. ten Broek and R. B. Wilson, “Public Assistance and Social Insurance — A Normative Evaluation,” *U. C. L. A. Law Rev.*, I (Apr. 1954), 251-62.

Performance norms, thirdly, must often represent choices on matters where conflict exists among various persons' economic freedom. The problem is not only that norms for prices and profits relate to people's income and wealth, but that approval or disapproval of almost any facet of competition — entry, cost minimization, innovation, etc. — implies approval of losses of various kinds of opportunities by competitors, suppliers, employees. It may appear that with certain kinds of performance — say minimum costs — the gainers could, income-wise, compensate the losers and still retain some benefit. Even so, it will not be feasible to identify the losers and their losses, to implement harmless taxes and bounties, or to compensate for losses which are a competitor's lot under capitalism. To know that compensation *could* be paid would settle nothing. A judgment on the redistribution is involved; the things payable as compensation may be valued higher by the losers than the gainers; and for the reverse choice the present losers might also be able to overcompensate. Even if compensation could not be paid, the choice might still be the best means to improve distribution. With compensation excluded, with no known set of enlightened and unchanging preferences held by a stabilized group of persons, and with no parity between events over time, formulating performance norms entails inelegant comparisons and subjective judgments of distributionally different nonoptimum alternatives.⁹

C. *A Presumption to Reconcile Interaim and Interpersonal Conflicts*

The subjectivity of resolving the second and third problems — interaim and interpersonal conflicts — seems in practice to raise surprisingly few qualms, or at least little explicit discussion.¹ The explanation probably is not merely that there is comfort in knowing that indecision and inaction on these matters would have the effect of a judgment too. More likely, the reason is that it is generally acknowledged that various interests of workers and investors must be subordinated, and likewise envy and other uncompensated effects of consumption, if market demands are to organize capitalistic production, consumer interests to be protected, and progress to be possible.

Accordingly, there arises the presumption that achievement of our materialistic focal aims will be greatest if production efforts and

9. Compare I. M. D. Little, *A Critique of Welfare Economics*, e.g., pp. 108, 122–25, 269.

1. To illustrate: If “the interests of society as a whole . . . conflict with the interests of persons who have assumed the risks of free enterprise . . . the latter should obviously give way.” Stocking and Watkins, *Monopoly and Free Enterprise*, *op. cit.*, p. 101.

conditions are the ones best likely increasingly to satisfy the *buyers'* desires which the community allows expression, ranked by the size of money demands. The advantages which follow from general compliance with this principle accrue to everyone at least as a consumer, and while specific exceptions should not be ruled out, they encounter the interests not merely of contemporaries, but of future generations. This seemingly modest rebuttable presumption greatly simplifies and facilitates the analysis. It enables us easily to condemn entry barricades and deceptive advertising and to condone innovation, economizing, reallocation, and profits tied to relative performance — although each poses interaim and interpersonal conflicts.

The conflicts, to be sure, are resolved for only the most obvious choices. At subtle points, separate subjective judgments are still essential, as in deciding that, say, zero is the level of economic profits departures from which should, and can, have positive justification. But a significant part of the subjective foundation of the performance norms can be reduced to merely an agreement about the named rebuttable presumption — an agreement that it operationally formulates the relevant part of the egalitarian and materialistically progressive aims of Western culture — together with an agreement that the whole question of workable competition is important and that truth and social welfare, rather than personal or other interest, should govern our solutions. Given agreement, what remains is often a factual inquiry.² The inquiry would be concerned in the case of performance norms to decide what state of the various performance dimensions would imply maximum service to buyers' desires, and in the case of structure and conduct norms, to decide what state of their dimensions would imply maximum effects on performance.

It is important to know that the same, and reasonably manageable, issues of values are involved however the workability criterion is approached. But this also implies that we must look elsewhere to decide which dimensions' being satisfactory should be necessary and sufficient for workability.

IV. NECESSARY VS. SUFFICIENT CONDITIONS

I have argued that the importance of structure and conduct attributes derives from their implications for performance. It follows that the conditions accepted as sufficient for workability ought either themselves to represent attainment of satisfactory performance or

2. Compare Meriam, "Bigness and . . . Competition," *op. cit.*, p. 125; J. W. Markham, "An Alternative Approach to the Concept of Workable Competition," *American Economic Review* XL (June 1950), 358.

else to be structure or conduct attributes which promise it. The latter approach is adopted by Adelman, Edwards, Mund, and Stigler, as the enumeration above indicates. The norms that they suggest are tests with which one can judge whether a market's structure or conduct is such as to produce favorable performance. This approach is the better one if, as Lewis argues, the essence of working and workable competition lies in compelling forces driving business to behave economically.³

A. Performance Cannot Be Inferred from Structure-Conduct

Unfortunately, fulfilment of structure or conduct criteria could not promise even usually to produce performance which is satisfactory. The idea is not tenable unless one defines "satisfactory performance" as any results which a law-abiding group of capitalists happens to produce; or else one (a) forgets many aspects of performance, (b) disregards for the aspects remaining the unfavorable side of certain effects produced on individual dimensions, and (c) puts unbounded faith in the occurrence of the favorable side of the effects which may be produced by the operation of certain market forces — price warfare, countervailing power, actual and potential entry, creative destruction, rivalry for its own sake, a long-range viewpoint, fear of prosecution, laxity toward stockholders' interests, distaste for profiteering, a sense of social responsibility, and a tendency of irreversible wage increases to discourage and disperse the taking of occasional large returns.

Most of these forces, however strong, are volitional and occasional. Room for managerial discretion is inherent, and with it the chance of choices that are unprofitable, or profitable but antisocial. Discretion inevitably accompanies free will, uncertainty, change, supramarginal status, immobilities, and indivisibilities. No structure conditions can compel firms to choose correct locations or co-operate on component standardization. Adding certain conduct prohibitions is clearly appropriate and could eliminate a few putatively antisocial practices; but entry faces natural barriers too, unfair tactics are not alone responsible for high costs or poor conservation, and the absence of coercion and collusion may leave firms free to cut prices or — most

3. B. W. Lewis, "Economic Implications of the Development of Large-Scale Organization," *Journal of Economic History*, XII (Fall 1952), 433, 436; untitled, *op. cit.*, p. 707. See also: Edwards, *Maintaining Competition*, *op. cit.*, pp. 10-11, 124; "Public Policy and Business Size," *op. cit.*, pp. 285-86. A. E. Kahn, "Standards for Antitrust Policy," *Harvard Law Review*, LXVII (Nov. 1953), 39. Mason, "Current Status of the Monopoly Problem," *op. cit.*, p. 1267. Stigler, *Testimony*, *op. cit.*, pp. 120, 129, 993, 996. Stocking and Watkins, *Monopoly and Free Enterprise*, *op. cit.*, pp. 94, 95, 414.

important — innovate, but does not assure that any will do so. To include in the conduct tests the requirement that competition should be aggressive, and yet attentive to social costs, would alter the conclusion only if someone were prepared to claim the knowledge and assume the prerogative of virtually prescribing each firm's every move. In a dynamic economy the aims, customs, rewards, and penalties which co-ordinate different persons' activities simply cannot closely determine their behavior. Indeed, market incentives are able to designate and evoke aspects of desirable behavior not merely when enterprises are constrained by conditions like substitutes' availability and opposites' knowledge, but when producers have freedom to innovate and latitude when successful.⁴

This is not to deny that it is often possible and sometimes appropriate to alter a market situation so that less scope exists for managerial discretion and social responsibility, if only by government competition. It is to say that no practicable set of structure or conduct requirements, and especially the incomplete set usually mentioned, can assure that performance will be satisfactory.⁵ Whether performance is favorable can be inferred only with data on performance. Consequently, if data on performance are available, fulfilment of structure-conduct norms should not be sufficient for workability.

B. A Purely Performance Criterion Is Inadequate

Their fulfilment should, nevertheless, be necessary. The reason why is related to two well-known difficulties of a purely performance approach.

The first is that it is often feasible to detect at best only extreme instances of good and bad performance.⁶ Cost performance can roughly be surmised from engineering, intraindustry, intertemporal, interindustry, and international comparisons; but even such light as

4. Compare: Bain, "Price and Production Policies," *op. cit.*, pp. 156-58; "Workable Competition in Oligopoly," *op. cit.*, p. 38. J. M. Clark, "Financing High-Level Employment," in P. T. Homan and F. Machlup (eds.), *Financing American Prosperity*, pp. 122-25; *Guideposts in Time of Change*, *op. cit.*, p. 62. Griffin, *Enterprise in a Free Society*, *op. cit.*, pp. 308-9, 333-34. Wallace, "Industrial Markets and Public Policy," *op. cit.*, pp. 109-10, 119n.; "Monopolistic Competition and Public Policy," *op. cit.*, pp. 274, 278.

5. For this reason and aside from the demerits of atomism, which are mentioned above and below, I must disagree with Mason that manyness plus no collusion is sufficient but not necessary for workability. See Mason, "Competition, Price Policy, and High Level Stability," *op. cit.*, pp. 29-31; *Economic Concentration*, *op. cit.*, pp. 36n., 38; and likewise Stigler, "Mergers and Antitrust Policy," *op. cit.*, pp. 181, 183. In addition, I argue below that no collusion — but certainly not manyness — *should* be necessary.

6. Compare Mason, "Current Status of the Monopoly Problem," *op. cit.*, pp. 1281-82.

these provide passes only dimly through a cloud of difficulties — variations in demands, products, product mixes, input prices, input qualities, long-lived commitments, cost accounting, obsolescence problems, distribution costs, etc. Likewise, in appraising profit performance, only certain extraordinary profit or loss rates could fail to be rationalized by the many accounting problems and the justifications of risk-bearing, innovation, cost performance, and resource reallocation.

The second difficulty is that many aspects of performance are not readily manipulable by public policy. Some, it is true, can be manipulated directly — usually aspects which are simultaneously dimensions of structure or conduct. Public policy can to some extent influence conservation, sleeping patents, dishonesty, advertising, adulteration, labor safety, and serious social costs. But various harmful effects would follow if public officials tried to enforce good results in costs, profits, volume, or progressiveness. Poor performance in these and other dimensions often can or must be attacked with remedies aimed at revising structure or conduct — say prohibiting a delivered-price system — in ways indicated by structure-conduct norms. In addition, structure-conduct revisions privately undertaken — like starting price guarantees — could be appraised by public policy only with the relevant structure-conduct norms.

While these two difficulties would make a performance criterion hard to apply, they do not in themselves justify adding structure-conduct requirements as necessary conditions of workability. However, they imply the reason for doing so. The reason is that it might be possible, by improving structure and conduct — say by requiring better information for buyers — to have even what appears to be good performance made either somewhat better or somewhat less the result of accident or discretion. Granting the point that we cannot conclude that performance is satisfactory by analyzing only structure and conduct, we could still be reasonably confident that certain structure-conduct revisions — perhaps compulsory grade labeling — would be contributors to, or prerequisites of, improved performance. Moreover, performance or our data on performance may reflect recent changes in the market's structure or conduct, like mergers, only after a significant lag. Analysis of the changes may give an indication that currently satisfactory performance will deteriorate or unsatisfactory performance correct itself. Consequently, before it is clear whether recent and prospective performance are satisfactory, reference must also be made to the market's structure and conduct.

The concepts of workability which have proposed criteria for

structure, or conduct, or performance, or combinations thereof, therefore, are not incompatible. Performance is what is of ultimate importance here, and whether performance is satisfactory can be judged only with performance criteria. But improvable, fortuitous, or passing performance can adequately be judged only by adding structure and conduct criteria. The conditions adopted as sufficient for workability will promise satisfactory performance only if they include dimension norms of structure, conduct, *and* performance.

C. The Monopoly Problem Encompasses All Market Dimensions

The dimensions of each stratum which most clearly need attention are indicated by the list of dimension norms presented earlier. The list is formidable. Even it, however, omits a number of significant market characteristics, as a glance at the appendix will show. Presumably the characteristics neglected have appeared less important, less susceptible to analysis, or less related to the problem of monopoly. Yet analytical simplification here is dangerous.

The danger is partly that an analysis which omits significant variables may nevertheless be taken to furnish reliable normative conclusions about the total state of affairs in a market, including the state of aspects like social costs which were neglected. More serious, the danger is that an analysis covering only the traditional variables may not furnish reliable conclusions even about many of them. Those listed intertwine and interact with others, so that what appears to be or to produce a relatively favorable condition in the usual dimensions — say concentration, independent profit maximizing, or volume of sales — may actually reflect or produce a highly unfavorable condition in directions ignored — say component standardization, helpfulness to other economic units, employee training, and factor availability. The sum total of the latter characteristics or effects may in some markets be crucial.

As a result, if the concept of workable competition is to provide a reliable criterion for judging whether a market situation is socially satisfactory, it must ignore no dimensions of normative significance and appraise simultaneously those which are interdependent. If this view is correct, then the "monopoly problem" being resolved here is a broad and interwoven complex which covers all variables that represent and influence the achievement of socially satisfactory performance.

It is misleading to pose the monopoly problem as the evils which may or do arise from the presence of a high degree of concentration of market supply, or from the collaboration or pricing or predation

which may accompany it — misleading, that is, even apart from the difficulties of market delineation or the fact that innovation sometimes produces at least temporary high concentration. It is misleading because the traditional argument behind this conception neglects many aspects of performance and also many aspects of structure and conduct which, interdependent with concentration, influence performance. Moreover, with low concentration, too, there exist problems of excessive or insufficient profits, high and rigid margins, excessive costs, heavy and deceptive advertising, complacency and caution — or actual suppression of innovations — etc.; and some greater degree of concentration might improve the market's over-all performance.

Once we abandon the presumption in favor of atomistic competition, the problem of monopoly makes sense only when it is viewed as nothing less than the whole problem of attaining a socially optimum situation in individual capitalistic markets. For this, all market characteristics are interdependent and pertinent. Referring then to those variables over which men have control, a market should be adjudged workable only if the state of *all* its malleable dimensions is satisfactory.⁷

D. Delineating the Investigation and the Market

A problem is presented by the fact that each dimension raises questions requiring a different scope of analysis. An appraisal of collaboration among the suppliers involves a different range of investigation than does appraisal of their conservation or their diversification economies.

The investigation does have limits. Specialized and pointed handling of the problems related to individual products obviously requires that issues be excluded which can be analyzed or controlled only in terms of sector- or economy-wide conditions or aggregates — issues like the correctness of inputs' prices or the appropriate ratios of investment to GNP.⁸ At the other extreme, it is clear that criteria of workability cannot be confined to matters evident in the financial

7. Included here is a satisfactory state of various aspects of an industry's labor relations. However, labor markets as such are ordinarily not equivalent to commodity markets for purposes of analysis or policy. See Edwards, *Maintaining Competition*, *op. cit.*, pp. 77-90.

8. Business cycle considerations need special comment. Reliable norms pertaining to a market's efficiency must presuppose that alternative employment is available for hired or released resources, that supply prices reflect real costs, that allocation considerations retain importance, that demand prices are comparable over time — in other words, that cycles have been conquered. The basis for the supposition can be a presumption that government will maintain continuously appropriate levels of aggregate demand. Or the basis can be a judgment

statements of individual firms. Such criteria would find different issues raised by the firm's various products and geographic areas, and also would neglect variables, like collaboration, which represent or result from the activities of a number of concerns. In between, however, it is difficult to say more than that investigation should extend outward from a convenient starting place as far as is necessary to resolve each of the issues in the market originally delineated.

The most useful criterion for the original delineation appears to be that of symmetrical high arc price-quantity cross-elasticities of demand.⁹ The idea is perhaps best expressed by saying that the workability criterion should formulate standards that apply to the *related* activities of all those who supply or demand all outputs *almost* any two of which are so akin that a *moderate ceteris paribus* reduction from the *prevailing* price of *either* would *soon* transfer to it a *substantial* proportion of the quantity *demand*ed of the other.

Even as a starting point, however, this criterion may, for convenience or want of knowledge, be supplemented or supplanted by a grouping based on similarity of producers' major inputs or processes. If so, the market's supply will not necessarily consist of only one product, and all the relevant volume of the products included may not be within the market. The original delineation is not critical. Conclusions can legitimately be reached — for instance that concentration is too low — which would apply *a fortiori* to a broader or narrower segment. What is critical is recognition that, once delineated, various aspects of a market's condition can be appraised only if investigation is extended to relations with outputs and enterprises which are linked not only by interchangeability to buyers or technological considerations, but also by complementarity, common sources of supply or demand or access thereto, common watersheds, etc.

My principal concern in this section has been to show that a satisfactory state of *all* malleable dimensions should be necessary and

that progress would unfortunately be stifled if the rules appropriate to stable full employment were violated during recession and inflation.

Paradoxically, various dimensions, like price flexibility and profits, ought simultaneously to have their state and time pattern judged partly by their contribution to the conquering of fluctuations — if criteria can ever be established. At least dampened cycles are in fact going to occur, and particular market conditions may aggravate or relieve them. Likewise, the norms for various other dimensions — business mortality, countervailing power, continuous price competition, potency of entry barriers, plant adaptability to a wide range of outputs, etc. — must recognize that their significance changes with the existence and the stage of cycles.

9. High cross-elasticities of supply are more conveniently acknowledged under the condition of entry (or the own-elasticity of supply for transference).

sufficient for a market's workability. What it means for a dimension to be "satisfactory" has necessarily been implied, but has yet to be elaborated. Herein lies the second fundamental point of difference among the eighteen writers identified earlier.

V. REMEDIABILITY VS. CATEGORICAL NORMS

Baum, Markham, Oxenfeldt, and Stocking and Watkins have suggested, in effect, that competition be considered workable if there are available to public policy no remedial measures which would improve the situation.¹ The suggestion is not enough to make these writers' views alike, since their criteria of improvability vary, as the itemization above indicated. But whatever dimensions are emphasized, the idea here is of fundamental importance. Except for certain attributes to be named below, each dimension norm should call for unimprovability of the dimension, the completion of all available beneficial changes — perhaps a dissolution in the market — rather than the attainment of some abstractly determined substantive condition — like manyness.

A. *The Place of Generalizations About Structure-Conduct Attributes*

Structure and conduct norms which attempt abstractly to name conditions favorable to good performance encounter serious problems.

1. Baum states, "If the 'probability' of more effective competition with increasing numbers (and suitable other policies) can be demonstrated in any particular case, and if this policy is preferable to other alternatives, then the new situation would be called 'workable' regardless of whether or not it conformed to a given structural standard of competition." *Workable Competition in Tobacco*, *op. cit.*, pp. 65-66. Markham writes, "An industry may be judged to be workably competitive when, after the structural characteristics of its market and the dynamic forces that shaped them have been thoroughly examined, there is no clearly indicated change that can be effected through public policy measures that would result in greater social gains than social losses." "Alternative Approach to Workable Competition," *loc. cit.*, p. 361; *Competition in the Rayon Industry*, *op. cit.*, p. 204. Oxenfeldt argues that existing performance should be compared with either "the best performance imaginable," as a theoretical matter, or with the best performance attainable in practice, in terms of political realities; all changes should be made "that would improve the over-all performance of the industry" after the costs of effecting such changes have been considered. *Economics for the Citizen* (New York, 1953), pp. 333, 347; *Industrial Pricing and Market Practices*, *op. cit.*, pp. 90, 559. Stocking and Watkins declare, "Competition is workable in the sense (1) that it is preferable to the best alternative 'competitive' arrangement practically attainable; (2) that such market control as sellers can exert is slight and, under the particular circumstances, does more good than harm." *Monopoly and Free Enterprise*, *op. cit.*, pp. 97-98. Compare also Mason: "The notion of workable competition . . . insists on the prime importance of an examination of possible remedies." Generally there is "Small reason for attacking an existing situation, if no better set of market relations can be put in its place." *Economic Concentration*, *op. cit.*, pp. 382, 383-84.

The categorical norms listed earlier — that the number of traders should be appreciable, quality differentials moderate, collusion and predation absent, etc. — are highly ambiguous and often presuppose a sharpness and accuracy in market delineation which is far beyond that required merely to choose an area for investigation. If in a particular context categorical norms do become explicit, the desirability of what many of the norms demand would be found to depend on the degree to which many others are or could be fulfilled. Whether a given degree of buyers' concentration is suitable depends on the characteristics of the sellers' concentration curve; sellers' concentration, on the condition of entry; ease of entry, on price flexibility; price flexibility, on quality standardization; quality variations, on buyers' information; buyers' information, on secret price concessions; price discrimination, on buyers' concentration.

In addition, the feasibility and desirability of many of the specific demands would depend on the relevance of a set of considerations which are both numerous and different from one market to the next. This is recognized by those who require for supply concentration only that the leading seller not exceed the share required for reasonably optimum real costs. Even this qualification is inadequate, however. It is not only costs relative to market share that vary from case to case, and not only cost performance that is influenced by concentration. A norm for concentration of supply is not helpful if it neglects the fact that markets are passing at various rates through various stages of development — some with good substitutes, some with patents — face different sorts and degrees of demand and cost variability, and have different possibilities of victory by one or four enterprises. Likewise, a norm for quality differentials is not helpful if it neglects the variously specialized character of different outputs; for collusion, if it neglects the need to approve — nay, demand — various types of co-operative action; for exclusion, if it neglects the economies which accompany certain forms of input and outlet pre-emption.

The source of the obstacles to reliable normative generalizations is not far to seek. Structure and conduct conditions should be judged by their implications for performance. But there is often no generic answer to the question of what effects on performance — say on the amount and quality of sales promotion — tend to be produced by a given state of a structure or conduct dimension — say by manyness; or to the question of what effects — variety-increasing or decreasing, for example — are needed; or to the question of whether the likely advantages available in one direction — perhaps 2 per cent lower costs this time — should be sacrificed for the available gain in another

— perhaps a diffusion of power among two more managements. Reliance on normative generalizations would often commit the gross error of prescribing aggravation of dimensions which even a casual observer could see are already seriously deficient.

These problems can be minimized by passing judgment on a market's structure and conduct only after referring to the circumstances prevailing in the particular case, the intent or design underlying the firms' actions, the situation in related markets, and the precise revisions that are expected to be an improvement. The norms for most dimensions of structure and conduct should demand in the abstract merely that any changes from the existing situation be completed which, in a given market, would tend in a net sense to promote performance, either directly or by effects on other facets of structure and conduct. The norms should consist of the abstract demand for favorable changes, plus a body of analysis which assembles the relevant descriptive generalizations and notes the circumstances which should turn a normative decision one way or the other in a particular case.

Ideally, such analysis would trace the net effects that various revisions — such as ending a price reporting scheme — would produce, in different types of situations, on every other dimension of structure, conduct, and — at least ultimately — performance. In fact, a priori analysis often can supply only a limited number of rebuttable presumptions and contingent tendencies. It might, for example, indicate that weakening price leadership should not impair product quality or smoke abatement; would tend to reduce price flexibility unless the weakening occasions aggressive maneuvers or better adjustments to changing demands and costs; and would tend (perhaps excessively) to reduce profits unless the leader maintains prices which otherwise in fact would move, and move up or down just to collectively more profitable levels.

Ideally, *ad hoc* investigation would remove the qualifications to such predictions. The interdependence and flux in market situations, however, insure that the effects actually produced will be, if not accidental, at least largely undiscernable with even the best combination of a priori and *ad hoc* analysis. Something glaring in a market's structure or history, like price wars, may indeed rebut the usual presumptions. But even the most careful investigation must often merely rely on what a priori seem to be the most likely results.²

2. Compare E. S. Mason, "Market Power and Business Conduct: Some Comments," *American Economic Review*, XLVI (May 1956), 475: "... since we are dealing with an economy in process of development, a judgment on the consequences of any particular part of it — say a combination . . . — can only be a historical judgment as these consequences 'unfold over decades' and a partial

Moreover, the individual desirability and the collective weighing of the apparent effects must often remain a matter of qualitative appraisal. Theory will often produce opposing desiderata. It might suggest that less flexible prices will engender simpler management, longer-lived catalogs, less elastic expectations, less speculation, better buyer planning, etc., but may also imply that prices are excessive from the beginning and fail periodically to adjust to substantial and protracted shifts in costs and demands. The appropriate compromise among such conflicting and variable gains must be a matter of personal judgment. At times there will be no alternative to resorting to a type of inference that is rejected by the nonperformance approaches to workability. The direction of reasoning sometimes must run from finding seemingly favorable performance — say in costs or product development — to concluding that existing structure and conduct — perhaps sellers' scale or concentration — are well suited to the peculiarities — perhaps the managerial situation — of the particular market.

Guesswork and personal appraisal clearly cannot be avoided in predicting and weighing effects. But the gross errors can be, by knowing the specific circumstances and relating the imperatives to the market's needs and possibilities.

B. The Categorical Norms

As indicated above, for some dimensions the abstract analysis may yield more definite conclusions. It may disclose that a situation should always be regarded as improvable if it presently includes or omits certain patterns or practices.

Abstract approval or disapproval can apply only to a pattern or practice which is defined concretely. Otherwise, the usual process of appraising its effects in a particular market would be required in order to place some condition within a category of events that have approved or disapproved abstractly. Abstract appraisal *should* apply only to conditions which, in the net, always tend to improve or worsen performance; or just usually so tend, in cases where the frequency and importance of exceptions is minor compared with the need for generalizations that, joined with legal prerequisites like intent, can be made into laws swiftly and easily enforced and understandable and compliant for businessmen.³ An isolated seller's judgment, since these repercussions reverberate throughout an economy. . . . An attempt to push enquiry into effects very far is clearly an invitation to non-enforcement."

3. But, as Brewster observes, if per se rules are clearer and more objective, they may also be less avoidable and less related to blameworthiness. My discussion, however, concerns Brewster's first question, standards of workable

exclusion of independent firms, to illustrate, would fail both tests.⁴ The concrete thing is the instrumentalities — perhaps scale economies, customer loyalties, or talent accumulations — by which exclusion appears as a result; and the instrumentalities may lie beyond control, or be specifically justified by their own performance implications, or be artificial entry barriers which originally justified investment in the project.

Even in the light of creative destruction, however, certain dimensions do pass the test for categorical norms. This conclusion could be supported only with extended discussion, since each dimension norm is a topic in itself. The candidates, however, are well known.⁵ At the structure level there are various interlockings and affiliations which almost automatically undermine the aggressiveness of otherwise independent competitors. At the conduct level, deceptive advertising and specific forms of unfair, exclusionary, predatory, and coercive tactics should be mentioned, along with resale-price maintenance and specific forms of restrictive overt collaboration.

Although few to enumerate in this form, these patterns and practices are of immense importance. To prohibit and attach a sentiment of wrong-doing to them is properly at the heart of antitrust policy; and it is only regrettable that, as argued earlier, their elimination is not sufficient for workable competition. But their absence is what it most obviously implied in saying that it should be a necessary condition of workability that no malleable dimensions of structure or conduct are improvable.

C. *The Nature of Performance Norms*

Unimprovability rather than specific attainments must usually be the keynote also of norms for performance, including dimensions which do not overlap structure and conduct. The same results — say zero profits, optimum costs, nonpersuasive advertising — cannot and, in view of their implications, should not be realized in markets growing and declining, fluid and immobile, innovating and static, expert and ignorant. Moreover, the individual desiderata conflict — product development and uniform profits, technique innovation and uniform efficiency, variety and mass production. Performance norms, consequently, must be sufficiently general to apply to markets in

competition, not his second, standards of fairly and feasibly enforceable competition. See K. Brewster, "Enforceable Competition: Unruly Reason or Reasonable Rules," *American Economic Review*, XLVI (May 1956), 484–86, 488.

4. Contrast Edwards, "Public Policy and Business Size," *op. cit.*, pp. 285–86.

5. See especially the thoughtful appraisal by Edwards, *Maintaining Competition*, *op. cit.*, *passim*.

different circumstances; the norms must acquiesce in the necessary effects of structure-conduct conditions which, however unfortunate, are not remediable even potentially, like immobilities,⁶ and they must embody concessions to one another to form a mutually consistent "interrelated complex."⁷ The norm for economic profits, for example, should not call for profits to equal one universally appropriate rate, but rather to vary up and down from zero an amount which, in view of its life expectancy, acknowledges the risks, innovations, cost performance, and mandates for reallocation of individual members of individual markets.

This is not to say that our standards should be watered down to some notion of a tolerable level of performance. In practice it would be feasible to detect and worthwhile to remedy only extreme cases of bad performance. But it would be quite arbitrary to lay down a rule which decides abstractly whether various remediable deficiencies fall within limits of tolerance. A performance dimension should not be considered satisfactory unless it is as favorable as "unavoidable physical and economic characteristics" permit.⁸ This approach refuses to be content with performance which, though good, could be better. Also it often makes it unnecessary to know what ideal performance would be, but only that actual performance is improvable — perhaps that costs have been raised even one dollar by sinecures or featherbedding.

Of course, what circumstances are unavoidable and what they require — say lobbying, contributions, bonuses, thick carpets, stand-by capacity — are matters of degree and always subject to dispute. In addition, it is clear that a policy instrument for a capitalistic economy must make some debatable concessions to the motives, tendencies, and institutions of capitalism — for example, acquiescence in acquisitiveness and property incomes; and that especially good reasons must be offered for what amount to recommendations to move away from the prevailing state of capitalistic arrangements — for example, subsidies, utility-type regulation, entry by state enterprises. In fact, contemporary analysis inevitably will presuppose a large part of the prevailing cultural and physical environment. The real danger here probably lies less in failing to recognize what is practically unavoid-

6. "... there are some severe limitations, arising out of existing technology, spatial considerations, population distribution, and basic institutional conditions, to a rearrangement of market structure and business practices." Mason, *Economic Concentration*, *op. cit.*, p. 382.

7. Bain, "Price and Production Policies," *op. cit.*, p. 167. See also, Wallace, "Industrial Markets and Public Policy," *op. cit.*, pp. 126-27.

8. Clark, "Orientation of Antitrust Policy," *op. cit.*, p. 98.

able than in forgetting that conclusions derived from apparent essentials will have validity only in relation to the existing stage of development, and will need to change with circumstances in order to promote a given set of goals.⁹

There is also a danger that performance norms will be biased in favor of the status quo in order to avoid difficult questions. As much as possible, the norms should not neglect social or nonpecuniary gains and losses by assigning status only to market-type wants and marketable resources. Nor should they view the potential of practices like standardization to be limited to what autonomous firms would achieve. Nor refuse — on principle, not budget — to appraise the consistency, adequacy, or necessity of whatever input prices, worker exertions, financial manipulations, product durabilities, demand fragmentations, etc., happen currently to prevail in the market. To incorporate these biases would indeed simplify analysis and control. But it would often do so at the price of confining attention merely to whether firms succeeded in maximizing their profits.

Correspondingly, no attempt should be made here to separate composite performance effects — say buyer irrationality, excess capacity, concentration of financial power — into those for which producers could be “blamed” and those which lie beyond their purview, knowledge, incentives, or power. Cause and blame properly enter when and if the question arises of finding a successful and legally appropriate remedy. For purposes of normative appraisal, the unimprovability that performance norms should demand should mean all that unavoidable circumstances permit.¹

D. Remediability vs. Potential Remediability

The availability of remedies is a question that cannot be *wholly* postponed in judging improvability. If a dimension is improvable in any useful sense, it is not only substantively deficient, but also procedurally correctable. There is little avail in condemning conditions that could be corrected only if it were feasible to prohibit all foolish purchases, to prevent any businessmen’s intercourse, or to dissolve every oversized concern. Moreover, whether some revision — say a dissolution — would represent an appropriate change may depend on

9. Compare Edwards, *Maintaining Competition*, *op. cit.*, p. viii.

1. Compare: Clark, *Guideposts in Time of Change*, *op. cit.*, p. 50 n.; “Orientation of Antitrust Policy,” *op. cit.*, p. 98; *Preface to Social Economics*, p. 44: “Toward a Concept of Workable Competition,” *op. cit.*, p. 243. T. J. Kreps, *Measurement of the Social Performance of Business* (Monograph 7, U.S. Temporary National Economic Committee, Washington, 1940), p. 43. Lewis, “Economic Implications of Large Scale,” *op. cit.*, p. 433. Mason, *Economic Concentration*, *op. cit.*, p. 383.

what other remedies are available. These might accomplish the effect sought — say lower concentration or lower profits — in better ways — say by easing entry or combining buyers — or might stand to be undertaken in any event and cause the initial possibility then to be inappropriate or to overdo its point. But the question of remedies should be postponed in one important sense.

A dimension norm should demand that its dimension be, not merely irremediable, but not even *potentially* remediable; that is, that it be in a state which, in view of economic and legal *necessities*, has no feasible and preferable alternative, or which is at least adequately self-correcting. This stronger requirement recognizes that a market situation which is potentially remediable may yet be irremediable because of obstructions or deficiencies in the prevailing public opinion, political climate, legislative action, or judicial interpretation. It would be deceptive for a situation — say wherein buyers' ignorance is causing quality deterioration — to be called "workable" just because obstructions, *which themselves are potentially remediable*, are causing an evident remedy — say publication of Bureau of Standards' product ratings — to be not immediately available, or cheaply enough, or without prejudice to better changes later. Procedurally as well as substantively, "unimprovable" should mean as favorable as unavoidable circumstances will permit.² The economic problems which are manifest and soluble in single-market performance, conduct, and structure dimensions should be held up to standards which, aided by the special label "unworkable," call attention to conditions that could in fact be improved by public policy, if only after agitation.

A dimension will be "unsatisfactory," then, unless all reasonably possible favorable revisions have been made, so that its present state is unimprovable — not remediable even potentially. A decision involves the following: (1) deciding that at least one change in the condition of the dimension might represent or produce better performance; (2) deciding that at least one remedy is reasonably available to effect the change; (3) predicting the direction and strength of the various effects of the change; (4) predicting the effects of employing the remedy itself; (5) deciding, on the basis of the surrounding circumstances, the other remedies being undertaken, the prevailing condition of the dimensions that would be affected, and the norms for those dimensions, which of the various individual effects would be desirable and which — perhaps because excessive — undesirable; (6) deciding that the desirable effects outweigh the undesirable; (7) deciding that there

2. Contrast Markham, "Alternative Approach to Workable Competition," *op. cit.*, p. 361.

are not better ways to obtain the desirable effects, including self-correction by automatic market processes.

A more concrete demand should be made abstractly only for the few, though important, structure-conduct attributes which can be defined without reference to their effects and which nearly always tend to enhance or worsen net performance. Whether other governable structure-conduct attributes would be desirable depends on what performance effects would follow, on what effects are needed, and on how much would be sacrificed of one effect and gained of another; these performance implications in turn depend on the degree to which various other characteristics are or could be present and on a variety of circumstances which vary from one market to the next. To pass judgment on such dimensions without relating a body of contingent generalizations to the particular market situation, and the particular improvements and remedies expected therein, would not merely involve guesswork, which cannot be avoided, but invite gross errors, which can be. For performance dimensions, the individual norms should be satisfied with neither more nor less than the full social benefit that economic and political necessities permit without excessive sacrifice in other performance directions. The appendix is one suggestion of how these prescriptions might be translated into specific norms.

The preceding discussion indicates what should be required for a dimension to be satisfactory. And, as argued earlier, if all malleable dimensions are satisfactory, a market should be judged satisfactory. But certain questions arise in aggregating the dimension norms.

VI. THE WORKABILITY CRITERION REFORMULATED

A. Optimality and Workability

The workability criterion is properly a simple aggregate of the exacting dimension norms advocated above. Such a concept, joined with an emotive label like "optimum competition," will stress the goal of having the best attainable state of affairs in individual markets and emphasize that public service is greater the closer is a market to its optimum. A market will then be adjudged "optimal" if and only if its actual performance seems as favorable in all respects as unavoidable circumstances permit, and its future performance apparently could not be improved or assured somewhat by a revision in any malleable dimension of structure or conduct.

The concept of "optimum competition" for each market is a useful target; but, alone, it would place most or all actual markets in the "suboptimum" category and fail to distinguish different degrees

of deficiency. The target concept should be complemented with a less rigorous standard, whose fulfilment could be said to represent public service which is "adequate" — sufficiently good that one should be neither excited nor inert about the deficiencies. The tone of "workable competition" fits here. A market might be considered "unworkable" only if it had "An extremely bad rating in any one direction or moderately bad or suspicious ratings in several."³ These ratings would be subjective, made in the light of the different importance of the several dimensions.

The workable-or-unworkable dichotomy, however, still fails to distinguish different degrees of under- or overfulfilment of the minimum requirements. Actual cases might be differentiated by ordering the possible combinations of success and failure in the various dimensions.⁴ Thus, a higher rank might be assigned to a "bad" rating only for concentration than to a "bad" rating only for profits. This procedure, however, would involve its own either-or distinction in the "good" and "bad" ratings; it would neglect to set off a very "good" in a highly important dimension — say progressiveness — against a barely "bad" in one with far less to offer — say transitory adjustments of prices; and it would produce a complex ranking system with only dubious significance.

A preferable alternative is to acknowledge that anything less than ideal is a matter of degree for both individual dimensions and the market situation as a whole, to grant that rating and weighting the several dimensions is subjective, and then to distinguish actual cases by gradations of suboptimality. This would involve a series of categories, including "workable," "almost workable," and "largely unworkable." Such labels should readily be abandoned, however, if the object is merely a fuller description of a suboptimum situation. Here there is no substitute for specifying the market's deficiencies in greater or lesser detail and with as little subjectivity and arbitrariness as possible. For instance, "optimum except that unit costs of one dollar are 50 per cent too large."

Of course, particular deficiencies in a market situation — perhaps mergers or advertising — can and often should be evaluated without undertaking to judge optimality in all other respects. The point is that to judge whether a situation *in toto* is as socially satisfactory as it could be, all governable market dimensions should, to the extent feasible, be weighed for reasonably practicable improve-

3. Bain, "Workable Competition in Oligopoly," *op. cit.*, pp. 37-38.

4. This is the approach of A. G. Papandreou and J. T. Wheeler, *Competition and Its Regulation* (New York, 1954), pp. 187-206.

ments; and in evaluating a possible revision in one respect — individually or as part of a total appraisal — effects on no other dimensions should be ignored without at least checking whether the facts of the case belie any usual simplifications.

The suggested criterion represents a denial that market situations can adequately be appraised by a priori standards which are confined to a few dimensions, which demand specific attainments, which judge improvability without considering possible remedies, which recognize only the stereotyped remedies, or which fail to heed a revision's effects in all significant directions. Such standards would provide a simpler and more objective means of obtaining answers, but answers often grossly misleading. Equally the criterion denies that the variety of market situations, all complex and in flux, can be appraised by purely *ad hoc* analysis. A certain few dimensions should be appraised with categorical norms; the remainder, approached with the rebuttable presumptions, contingent predictions, and qualified judgments supplied by accumulated economic theory. This mixture of theory and *ad hoc* analysis certainly yields no definite standard. But the very indefiniteness of potential remediability is exactly what permits a single criterion to be applied to the variety of different kinds of market situations. Something more concrete could apply to only a limited range of cases, and to distinguish them would require the same inquiry into the improvements possible in particular markets.⁵

B. Workability Is Relative to Various Subjective Judgments

A "workable" rating is relative, of course, to the variety of subjective judgments of value and fact which were encountered above and may here be summarized. (1) The investigator must decide that the whole question of workable competition is important and that truth and social welfare, rather than personal or other interest, should govern his appraisal. (2) He must decide that a certain segment of economic affairs can usefully be treated as a unit for purposes of normative appraisal. (3) He must have in mind a generally acceptable set of focal aims, whose meaning is sufficiently clear that it is possible to formulate performance desiderata in the face of interaim and interpersonal conflicts, and to decide which sacrifices in individual performance dimensions are just worth making in order to obtain concomitant advantages in other performance directions — conclu-

5. Especially the criterion does not demand (nor *necessarily* repudiate) attainments — say $P = MC$ — whose desirability presupposes that conditions elsewhere are also optimum in either sense. It presupposes merely that enough is known about other markets to reach sensible conclusions about the immediate situation where it is interdependent — as a complement, substitute, further processor, etc.

sions reached sometimes by stressing buyers' sovereignty, sometimes by making exceptions in the interests of workers, owners, and external diseconomies, and sometimes by making rather subtle choices among performance alternatives. (4) He must predict the direction and strength of various influences running among items in the categories of aims, performance, conduct, structure, and remedies. (5) He must decide what market characteristics or effects bear so directly on aims for a good economy that they should be regarded as dimensions of performance. (6) He must decide which of the effects that individual structure or conduct attributes produce on performance are desirable and which undesirable, and decide which of the former should be sacrificed and which of the latter tolerated in order to have the effects that in the net most enhance performance. (7) He must decide what likelihood is sufficient to warrant categorical approval or condemnation of conditions which just usually improve or worsen net performance. (8) He must decide what limitations are set by economic and political necessities on the achievements and remedies that can be expected in a given market, including the concessions that should be made to the capitalistic nature of the environment.⁶ (9) He must decide how to rate and weight various individual deficiencies in order to obtain a composite rating for a suboptimum situation.

Despite the part of this subjective foundation which pertains to social goals, the concepts of optimum and workable competition have limited implications for public policy.⁷

6. The workability criterion could apply to socialized or "regulated" markets only if the categorical dimension norms were modified and the supporting analysis of the others made to cover for all cases the circumstances which should lead to different normative conclusions. Furthermore, an "optimum" rating does not imply that socialization or utility-type regulation would be unable better to serve the aims from which the collected dimension norms are derived. This would follow only if any special concessions and requirements made in appraising items like profits and losses and social costs and benefits were weighed against results under other kinds of arrangements.

7. "Competition" may be a misleading term here. Compare Oxenfeldt, *Industrial Pricing and Market Practices*, *op. cit.*, pp. 89-91. Also, see Mason's distinction between workable competition and effective business performance in "Current Status of the Monopoly Problem," *op. cit.*, pp. 1280-81. It is strange that a market containing a single seller or buyer, and conceivably one that is "regulated," co-operative, or nationalized, could at least for a time be judged optimally or workably competitive. Of course, in an important sense all suppliers provide options which compete for a place in buyers' budgets. The real explanation, however, is that "workable competition" has become the accepted expression for labeling emotively whatever standards a writer puts into the workability criterion, and the ones advocated here do not correspond to any of the traditional market categories. It would be better to speak just of "optimum" and "workable" market situations. It might then be unnecessary to emphasize that atomism and other approximations to perfect competition are neither necessary nor sufficient for workability.

C. Lack of Antitrust Implications

The collected dimension norms and their supporting analysis — including the distinction between categorical and potentially remediable defects — pose an ideal for market conditions and a basis for appraising dynamic capitalism. They suggest for what antitrust and other legislation might be enacted, how it might be interpreted, where cases might be selected for administrative and judicial action, what remedies might be applied, and how the success of public policy might be judged. The goals giving rise to the norms could be served by attaching the influence of law and public opinion to the concept of workability and therefore the opportunities that exist for remedial action.

But it should be clear that the suggested concept of workability names only an "economic" ideal, not anything that could or should be legally imposed. Limited time and resources and the constant change in market conditions make it obviously impossible to check every malleable market dimension for possible improvements. It is already a large task to enforce less than all of even the widely acknowledged substantive norms mentioned earlier, plus a sort of remediability check on high or growing concentration and certain other variables.

Moreover, "noneconomic" considerations narrow the application of such investigation as is feasible. Regulation which does too great violence to certain principles becomes inoperative, not worthwhile, and even self-defeating. Legal implementation should as much as possible not require detailed official knowledge; intrusions on privacy; surveillance and supervision; uncertain compliance; discretion concerning defendants, guilt, and burdens; prolonged and costly litigation; unequal treatment for persons similarly situated; penalties without fault, antisociality, forewarning, or avoidability; and impairment of performance-favorable motivation. Yet — to mention only one conflict — "unworkable" does not imply that cause or blame for deficiencies is attributable to the enterprises concerned.

The suggested concept of workability, therefore, is by no means a proposal that public officials should continually inspect private enterprises and, with broad prediction, qualitative appraisal, and personal judgment, either instruct managements how to give better performance or penalize them for failure to do so. If anything, the emphasis placed on incremental performance effects has less suggestion of interference with freedom of enterprise than would a comprehensive and meaningful set of norms demanding specific market patterns and practices. Legal implementation in fact should reflect some nice compromise between utilizing the opportunities for reme-

dial action, economizing use of the state, and acknowledging the importance of "noneconomic" considerations. Concrete policy recommendations would each be an extended topic, and if the Report of the Attorney General's Committee is precedent, may be left in other chapters.⁸

In any case, there is certainly no prospect of a solution to the monopoly problem in the sense that antitrust and other public policy — preventive and corrective — together with automatic market processes, will guarantee to produce workable market conditions. Wise public policies can, and in the United States do, help to a significant extent; but the factors mentioned clearly set limits on even their potential. If many markets are workable, the reason will be that the public has helped to create an environment favorable to the frequent operation of forces which remain volitional and occasional — altruism, rivalry, entry, countervailing power, and creative destruction.

APPENDIX

DIMENSIONS AND NORMS OF PERFORMANCE, CONDUCT, AND STRUCTURE

A catalog of dimensions and dimension norms supplies a generalized basis for the study of particular markets. It should cover every variable which in particular cases might prove to be significant.

8. Brewster observes that "the net goodness or badness of effects is not a proper subject of judicial determination." *Op. cit.*, p. 487. One may agree that such is undesirable in itself and for the effects which Brewster describes. This is inconclusive, however. Concomitant benefits are also important, though perhaps less evident than with regulating utilities. Unless someone (a nonjudicial FTC?) evaluates certain effects, many patterns and practices will go unchecked merely because they do not qualify for substantive norms and per se rules. There is often no other way to appraise combination, informational activities, standardization, delivered pricing, discrimination, leasing, patent licensing or pooling, tying arrangements, etc. Compare E. S. Mason, "Market Power and Business Conduct: Some Comments," *op. cit.*, pp. 474–81.

Kahn argues that practicability demands that antitrust standards relate to conduct dimensions; in particular, to collusion, combination, unfairness, exclusion, coercion, and predation. A. E. Kahn, "Standards for Antitrust Policy," *op. cit.*, pp. 30, 42, 50, 51, 54. With a few exceptions, like interlockings, I would agree that prohibiting certain conduct is the proper focus of antitrust policy in a narrow sense. However, I would include at least resale-price maintenance and discrimination as other points of attack, and add that structure and performance standards may be useful at least in selecting defendants and choosing remedies. Compare Edwards, *Big Business*, *op. cit.*, pp. 114–18; Mason, *Economic Concentration*, *op. cit.*, p. 383. Moreover, various aspects of structure and performance are good candidates for influence by more positive kinds of public policy — conservation, patents, imports, entry, promotion, labeling, product testing, grading, labor relations, financing, relations between owners and management, etc.

The present catalog is only tentative. It adds meaning to my prescriptions for dimension norms, but the meaning is only suggestive and the prescriptions stand independently.

Each of the norms offered here needs extended elaboration and support.⁹ The very enumeration of performance dimensions contains a subjective element. The catalog embodies a compromise between mentioning every difference among markets and referring to only a few oversimplified variables. It calls certain items "conduct" which from a different angle are structure, and some "structure" which contain elements of active conduct. It lists as separate dimensions items which obviously overlap, so that various dimensions shade into others within and among the several categories with merely differences of emphasis. And it makes a distinction which is largely arbitrary between two kinds of structure dimensions. The "malleable" dimensions are those which public policy might reasonably influence or rectify. For these the theory of workable competition should develop normative standards. The "intrinsic" dimensions are patterns determined by nature, technology, rigid institutions, past events, or chance. For these, not norms, but recognition of effects is appropriate — and some have effects dominant enough to rank the items high in classifications with other purposes.

I. PERFORMANCE DIMENSIONS

A. Ratios of economic profits to invested assets and net worth should vary up and down from zero an amount which, in view of their life expectancy, corresponds to enterprises' risks, innovations, cost performance, and mandates for reallocation.

B. The physical volume of sales should be the quantities that customers demand at prices which generate no continuing excess demand or inventory accumulation, which are not troublesomely oversensitive relative to adjustment speeds, which could not profitably be lowered, and which bear a relation to average total costs that generates appropriate profits — or a reasonable relation to ordinary marginal costs when the latter regularly differ in magnitude from average costs. Except where benefits are substantially unconfined, collection expenses are exorbitant, or public amenities are at stake, indivisible expansion should occur if total revenues are expected actually to exceed total costs, and contraction if revenues are regularly unable legitimately to cover variable costs.

C. Opportunities to reduce costs, improve on products, and build up markets should not be neglected, obstructed, or suppressed

9. See my dissertation, *op. cit.*, chaps. 4-6.

without good reason; obsolete or undesirable market conditions should be corrected with reasonable speed, directness, and equity.

D. Total money costs over the relevant time span and range of firms and products should be as close to the minimum legitimately required to deliver the use-values being delivered as is permitted by prior input commitments, reasonable delays, chances of further improvements, indivisibilities, and the inherent uncertainty and variability of demands and prices.

E. Sales promotion should have the quantity and quality that a responsible consumers' co-operative would choose in order to inform or otherwise benefit customers; buyers should have the opportunity to be well informed and trustful and should act rationally in their own interests.

F. The quality, variety, standardization, and location of sale of products should be curtailed only to effect the compromise between economy and individuality which informed customers would choose by an appropriate direct ballot.

G. Individual sellers and buyers should be law-abiding, reasonably helpful to related economic units and community organizations, and mindful of deleterious influences on public policy and international relations. Inattention to uncompensated social costs and benefits should not significantly distort allocation or distribution, create serious unnecessary costs, lead to the neglect of important activities, impede the minimization of depressions and inflations, or impair contribution to the national security in time of peace or war.

H. Producers should exert whatever favorable influence is needed and reasonably available on the quality, availability, training, and absolute and relative prices of inputs; producers should be effective in the discovery, extraction, management, deferral, imitation, and consumption of exhaustible natural resources.

I. Labor-management relations should not be characterized by ill-will, frequent shutdowns, featherbedding, exploitation, or wage increases to the neglect of price reductions. Workers should have protection against indignity, insecurity, arbitrariness, favoritism, partisan pressure, undesirable working conditions, and corrupt and undemocratic unions.

J. Management should have adequate power, sufficient incentive, and a motivation to strive for the most profitable long-term return on equity, except where such would obviously conflict with social welfare; simultaneously, managers should be responsible to holders of the enterprises' securities through procedures which do

not make a nuisance of owners' intervention, but nevertheless shortly replace unsuitable personnel and rectify improper practices, including practices pertaining to the distribution of profits.

K. Concentrated economic and financial power should have positive justification, and protections should exist against its abuse — e.g., by discrimination — vis-à-vis suppliers, competitors, customers, communities, and democratic liberties. The opportunity to establish a new concern should not needlessly be obstructed, and companies' survival should correspond to their comparative public service.

II. CONDUCT DIMENSIONS

A. Enterprises should not — individually or collectively, overtly or by threat — employ against actual or potential competitors, suppliers, or customers any of a variety of tactics whose means, motives, and effects disconnect the acting or suffering firms' survival from public service and classify the tactics as unfair, exclusionary, predatory, or coercive: deception, fraud, malicious interference, legal harassment, squeezing of supplier- or customer-competitors, sacrificially undercutting smaller or less diversified concerns, and pre-emption of inputs or outlets by needless acquisition or tying arrangements.

B. All forms of collaboration should occur among and between suppliers, customers, and others which promote their or others' welfare, except forms which clearly — if only mediately — tend, via profit and other performance, to damage interests beyond those that must be subordinated to achieving efficiency and progress. It is implied here that explicit and tacit collaboration, and extreme but still noncollusive leadership, should not have certain effects — dividing the demand or supply, pooling sales or profits, fixing prices, restricting output or purchases, coercing other concerns, or excluding competitors, suppliers, customers, or innovations — except when these are unavoidable aspects of arrangements which promise a net gain in performance, or when sellers or buyers co-operate in good faith to reduce prices that joint demand situations or irremediable formal obstacles would otherwise prevent reducing.

C. Minimum and maximum prices at which buyers may resell should not be more than suggested.

D. Economic discrimination in various selling terms should not prey on weaker sellers, disadvantage small buyers, cement collusive price structures, divorce buyers' pecuniary savings from real economies, rigidify channels of distribution, or refuse higher-priced demands while satisfying lower, unless there is adequate justification in favorable considerations — the good faith meeting of competi-

tion, changes in articles' marketability, processes of price competition and countervailing power, sampling of sections of the demand, stabilization or expansion of volume, equitable adjustment to ability to pay, or allowing production to occur at a profit. Refusals to deal should not be part of a group boycott or effectively made conditional on what may not legitimately be imposed.

E. Methods of determining terms of sales and means of effecting transactions — including standard contracts, rate books, price lining, discounts, credit extension, quality guarantees, consignment selling, leasing contracts, traveling salesmen, guarantees against price reductions — should be the ones most convenient and conducive to favorable performance in the market, and not to exclusion, conscious parallelism, or price leadership. Future prices should be announced or multiple-order contracts concluded only if the less concentrated side gains more in foresight or economy than it loses in defensive pricing or entry barriers.

F. Profit seeking and aggressiveness in rivalry and bargaining should be stronger or weaker than they are if this would tend to improve performance. The pattern in which buyers spread their individual patronage among sellers, and sellers spread their volume among buyers, should be altered if such would better respond to superior terms or would tend to undermine collusive market divisions, entry-blocking loyalties, or efforts to shield relatively inefficient competitors, suppliers, or customers.

G. Buyers, sellers, and potential entrants should be more receptive to change and should respond to various stimuli more promptly and rationally when such would tend to improve performance. Buyers should not without good reason either have elastic price expectations or judge quality by price or other superficial indicators.

H. Unless greater or lesser price flexibility has positive justification, sellers should disregard insignificant or ephemeral changes in demand and cost conditions, but respond to fundamental changes with conclusive revisions in their charges or markups.

I. Companies should partly or wholly combine or dissolve when such is to their owners' advantage, but only if this does not change the structure of any market in contravention of the norms for concentration, integration, diversification, or condition of entry.

III. MALLEABLE DIMENSIONS OF STRUCTURE

A. Interlocks can be created by holding companies, ownership, owners, credit, creditors, officers — including relatives and associates, reciprocity, patent licenses, technological pools, joint ventures;

these should be disclosed and altered if they unjustifiably threaten efforts of competitors to gain an advantage, raise questions of conflict of interest, permit exclusion of rival suppliers or customers, concentrate economic power, or involve trading on dangerously thin equities.

B. Sellers' and buyers' concentration curves should be different than they are if securing or avoiding entry, exit, combination, dissolution, or reapportionment would tend in the net to improve performance. Virtually every aspect of performance can be affected, directly and also via effects on collaboration, price variability, coercion, unfairness, heterogeneity, countervailing power. The effects involved include those of the remedies employed. The latter can consist of influence on unfairness, predation, sales promotion, survival, price fixing, entry barriers, exit difficulties, combinations, promoters, taxation, divestiture, and government competition.

C. Turnover among buyers or sellers should be altered if it fails to correspond to natural human attrition and a continual process of weeding out the truly less dynamic or efficient organizations. Weeding should occur at a pace slow enough to allow reabsorption of released resources and to avoid spreading bankruptcy from the inefficient to their particular suppliers and creditors.

D. Bigness, numbers of plants, markets, varieties, and products, patterns of balanced or disproportionate integration, and the number of points of contact among and between buyers and sellers should be altered if they represent, not a search for growth, economies, or assurances that are conducive to favorable performance, but a needless basis for financial power, predatory undercutting, suppression of interindustry competition where entry is blocked, squeezing, profit shifting, or mutual forbearance from aggressive action.

E. Trade, marketing, chamber, and professional associations should be created or reorganized to facilitate beneficial and eliminate harmful forms of collaboration.

F. Standardization, grading, and informative labeling should be employed or altered if such would tend to improve selection, reward quality, avoid a multiplication of meaningless alternatives, or reduce costs of production, exchange, maintenance, and usage — without imposing on variations in taste or obstructing progress in products and techniques.

G. The amount, speed, accuracy, and similarity of market and technical information should be altered if a substantial proportion of buyers, stockholders, or potential entrants lack the opportunity to be reasonably well informed from impartial sources, if traders

know too little about their costs, prospective demands, or other market features to behave economically, or if traders know too much about competitors' positions, plans, or trade secrets to behave aggressively.

H. Entry should be easier or harder than it is if more or less actual and potential entry would tend in the net to improve performance. Direct effects may occur on profits, volume, costs, progress, sales promotion, and discrimination. Indirect effects may occur via changes in collaboration, price variability, predation, coercion, concentration, heterogeneity, and youthfulness. Effects may also arise from the means used to alter potential entrants' cost or demand disadvantages or barriers to quickly seeking available profits. The means include influencing credit, training, apprenticeships, pecuniary economies, exclusionary and predatory tactics, sales promotion, market information, patents, government buying and selling, interjurisdictional trade barriers, legal requirements, and taxation.

I. Information, research, and other services by government, together with legal controls on prices, discrimination, production, entry, imports, quality, standardization, promotion, taxation, finance, intracorporate affairs, etc., should be altered if such would fairly and feasibly tend to improve performance.

IV. INTRINSIC DIMENSIONS OF STRUCTURE

A. Product Attributes.

1. Novelty, background of variations and likelihood of future changes, stage of growth.
2. Good vs. service.
3. Servicing requirements, advantages of unified sale and servicing.
4. Expensiveness.
5. Durability.
6. Divisibility.
7. Storeability.
8. Transportability.
9. Resaleability.
10. Complexity.
11. Imitability.
12. Physical change on sale (e.g., glass).

B. Production Characteristics.

1. Record and prospect of technological variations; channels of transmission of scientific developments.
2. Technological secrets.

3. Importance and costs of litigation.
4. Period, regularity, and predictability of production.
5. Output and input inventories.
6. Limitations on inputs' supplies, transferability, and readjustability.
7. Location, ownership, and method of acquisition of inputs.
8. Inputs' past and prospective price variations and speculation thereon.
9. Availability of various kinds of insurance.
10. Number, skills, turnover of employees; labor costs and methods of pay and promotion.
11. Union activities.
12. Use of nonhired inputs.
13. Sizes, proportions, and durations of fixed costs.
14. Costs of shutdown, maintenance, rehiring.
15. Joint costs, joint products, by-products.
16. Relations between costs and: utilization, scale, multiple plants, integration, and diversification.
17. Interfirm and interregional cost differences; equality of access to inputs.
18. Excess or stand-by capacity; ease and divisibility of expansion, contraction, relocation.
19. Ease of exit.

C. Marketing Features.

1. Producer vs. middleman vs. consumer vs. government vs. institutional uses.
2. Use-value to sellers.
3. Low vs. high income uses.
4. Replacement vs. new demand.
5. Luxury and economy models.
6. Importance to buyers of small quality differences.
7. Desires of buyers for constant, equal, or lower prices.
8. Separability of buyers and sellers.
9. Frequency of purchase and sale.
10. Variations in size of orders.
11. Produced for market vs. made to order.
12. Order-shipment lag.
13. Inherent output heterogeneity.
14. Importance of design, style, fashion, prestige, appearance, containers, credit, location, courtesy, personalities, trade names, and information furnished by sellers.

15. Availability of independent servicing and other complementary items.
16. Association of price and quality.
17. Customary or convenient prices.
18. Costs of price changes.
19. Price elasticities and cross-elasticities of demand, including intertemporal aspects and importance of substitutes and complements.
20. Pattern, record, and prospect of secular, cyclical, seasonal, and irregular movements in prices and volumes.
21. Sizes and overlapping of markets; spatial price patterns.
22. Proximity of buyers, of sellers, of each to other; stability of locations.
23. Friendships; codes of ethics.

D. Organizational Aspects.

1. Proprietorships, partnerships, syndicates, trusts, corporations, co-operatives, government enterprises.
2. Origins of owners and officers.
3. Connections between ownership and control.
4. Organizational structures; divisions; decentralization; accounting; channels of market and technical information.
5. Cash, credit, liability, and capital positions; sources and availability of finance for entry, expansion, innovation.
6. Time preferences, horizons.
7. Ages as organizations and members of the industry; methods of growth.
8. Positions in regard to taxes, subsidies, services, legislation, administrative regulation, litigation, public relations.

ECONOMIC STRUCTURE AND THE REGULATION OF TELEVISION

By HARVEY J. LEVIN*

I. Economic structure, 425. — II. Underutilization of the ultra high frequencies, 430. — III. Network practices and public policy, 439. — IV. Conclusion, 449.

Television's economic structure includes several major components: set owners, who view the programs and buy the products advertised thereon; stations, which broadcast or "clear" the programs supplied by national or specialized network companies, by independent program producers, or by the station itself; national, regional and local advertisers, who sponsor whole programs and buy time for interspersed commercial messages from the networks for clearance on chains of stations, or from the national "spot" representatives (for reaching certain specified markets only), and who ultimately support all other components; network organizations, which integrate stations into nationwide systems, selling the time of affiliated stations to network advertisers, supplying commercial and sustaining public service programs to the affiliates, and leasing from the American Telegraph and Telephone Company the coaxial cable and microwave facilities linking together the chains of stations that advertisers order.

We are concerned here with the stations and the networks, subject as they have been in recent years to searching Congressional scrutiny.¹ We shall examine evidence on the causes and consequences of and possible remedies for economic distress of stations operating in

* Hofstra College.

1. See *Television Inquiry, Hearings on S. Res. 13 and S. Res. 163 before the Senate Committee on Interstate and Foreign Commerce, 84th Congress, 2d Session; Monopoly Problems in Television, Hearings on the Television Industry before the House Antitrust Subcommittee, 84th Congress, 2d Session; and Status of UHF and Multiple Ownership of Television Stations, Hearings on S. 3095 before the Senate Committee on Interstate and Foreign Commerce, 83d Congress, 2d Session* (hereafter called *UHF Hearings*). Other useful documents in this series are: *Report of the Senate Commerce Committee on Television Network Practices* (hereafter called *Magnuson Report*), *Report of the House Antitrust Subcommittee on the Television Broadcasting Industry* (hereafter called *Celler Report*), *Television Network Regulation and the UHF Problem*, Memorandum prepared for the Senate Commerce Committee by H. M. Plotkin, 1955. Lastly see *Network Broadcasting*, Report of the Federal Communications Commission's Network Study Committee to the House Interstate and Foreign Commerce Committee, 1958 (hereafter called

the Ultra High Frequencies,² and the continued dominance of three national networks. We believe that the impact of network practices on the profitability of UHF stations, and the further tendency of UHF distress to implement the dominance of three national networks, must be analyzed carefully to formulate any effective program to revitalize UHF television, to promote a more competitive market structure, and to produce a more diversified program fare.

I. ECONOMIC STRUCTURE

But first a word more on five of the industry's main structural characteristics.

First, there are not only far fewer television than radio stations now operating but far less room for them, technically, in the broadcast spectrum, even if UHF channels are fully utilized. Television homes now receive programs from 510 commercial and 31 noncommercial stations, which contrasts with 3222 standard (AM) and 540 Frequency Modulation (FM) radio stations. Furthermore, the 1875 stations that can eventually operate in the 82 commercial television channels set a maximum that is several times below the maximum number of radio stations potentially able to operate in the 107 standard (AM) frequencies and the 80 commercial FM frequencies. Additional technical scarcities result from the fact that whereas the bulk of television channels and assigned outlets are in the Ultra High Frequencies, most operating stations are in the Very High Frequencies which, at best, can hold only 556 stations.

This brings us to our first area of inquiry; that of competition among stations for the set owners' attention, for network affiliation and for network, local and spot advertising business. In Section II

Network Study). This important study was released after the research for much of this article was launched.

2. This refers to that portion of the radio-TV spectrum between 300 and 3000 megacycles. The UHF band was first opened for commercial use in April 1952 after a thirty-six month period ("the freeze"), during which no permits whatever were granted to build more stations in the older Very High Frequency band (30-300 megacycles). During the interim, new engineering standards were formulated, wider spacing between stations was planned to minimize interference, alternative color television schemes were debated and tested, and the case for reserving special educational channels was more fully explored. New grants were postponed until these many fundamental policy issues were examined. But the 108 stations authorized to be built in the VHF band before September 1948 were permitted to complete construction and to operate. UHF and VHF channels actually occupy only a small portion of the broadcast spectrum which now actively uses portions of the range between 10 kilocycles and 10,000 megacycles. Standard (AM) radio operates between 540 and 1600 kilocycles; commercial and educational FM operates between 88 and 108 megacycles.

below we will see that stations operating in the Ultra High Frequencies can be heard by only one-sixth of the 42 million sets in use today, and that their signal strength and clarity are still deficient compared with those in the Very High Frequencies, which puts them at a great disadvantage in gaining choice network and advertising affiliations. This further acts to reduce their resources for the programming that might otherwise induce set owners to purchase all-channel receivers. When combined with the high costs of television operations and the basic technical restraints already noted, the additional shortage of comparable facilities resulting from the UHF distress makes network affiliation even more important than otherwise, and further weakens the bargaining power of affiliated stations. Under these conditions trade practices that operate to restrain competition require more than ordinary attention. Another problem, only touched in passing, arises from multiple station ownership, which upsets competition so far as multiple owners have more bargaining power than single station owners in gaining choice network affiliations and rate agreements. Because of the scarcity of TV outlets, moreover, multiple station ownership, like the UHF problem, operates to block the goals of local, diversified station ownership and program diversity.

The industry's next three characteristics are economic and institutional rather than technical: they concern the role of advertiser support, the place of networks in programming and finances, and the role of independent producers and spot advertisers.

*The industry is advertiser-supported.*³ Commercial stations derive most of their revenues from the sale of broadcasting facilities for designated time periods. National and regional advertisers buy station time and programs, or the right to intersperse commercial announcements in so-called "adjacencies," either from three national networks serving as sales agencies for nationwide chains of stations, or from thirty station representatives selling "spot" time in selected markets, on selected stations. Likewise, stations seeking sponsors and programs may deal either with the three national networks or with the independent producer-spot representative system. Some 350 advertisers bought network time in 1956, and some 4400 bought

3. Public and private outlays on noncommercial, educational stations are the only major nonadvertiser source of finance now existent, although another alternative is proposed by those who advocate subscriber services for commercial television. Considerable interest in this latter device did not prevent Congress from shelving action on it this past winter. (On subscription television see *Television Inquiry*, III, also FCC, *First Report on Docket No. 11279*, mimeo., pp. 57-1153).

national spot time, the top 50 accounting for 75 per cent and 45 per cent of purchases respectively.

Networks are crucial in station programming and finances. All but thirty-five television stations are now affiliated with one or more of three national networks which service them with programming and which sell their facilities to the network advertisers. Each national network organization operates a radio as well as a television network, owns and operates eight to ten radio-TV stations, serves as national spot representative for these stations and for some affiliates too, produces live and filmed programs, and has extensive connections with related industries like motion picture exhibition, or the manufacturing of radio-TV sets, tubes, transmitters, antennas, studio equipment, etc. Network-affiliate relations are governed by renewable contracts of from one to two years duration, which designate: the rate at which station time will be sold to network advertisers; the station's compensation for broadcasting network commercial programs; the number of nonsponsored programs, live and filmed, the network will provide gratis; the affiliate's "right of first call" to all network programs. The contracts further specify the "free" hours during which stations promise to carry network commercial programs without charge; and the affiliate's willingness to accept all commercial programs during certain "option periods" — consistent with the Federal Communications Commission's rules on the matter.⁴ In addition, the Columbia Broadcasting System and National Broadcasting Company place some fifty-five affiliates in the top markets on a "must-buy" list, and require all network advertisers to buy the use of their facilities for each program sponsored. Affiliates find their network affiliations valuable for at least three reasons: the superior popularity of many network programs, the compensation they receive for carrying them, and the expenses they are spared by not having to prepare their own programs, or to secure syndicated film strips to fill in otherwise unfilled portions of the broadcast day.

Independent program producers and national spot advertisers are

4. The FCC defines "option" as any agreement preventing network affiliates from scheduling programs before the network agrees to utilize a certain time period, or which requires a station to clear time already scheduled. Under existing Chain Rules, stations must retain the right to reject network programs found unsuitable, and may not bar or displace any network or grant clearances in option time without at least 56 days' notice (though they may legally displace spot programs without notice). Aside from option privileges, moreover, no station may enter into exclusive affiliations with a single network (though the category of primary affiliate approaches this); nor may it secure exclusive rights to any network's programs in a given area (although other stations desiring access to the program may be thwarted where the affiliate exercises its right of first call).

another source of programming and revenues for all stations, an alternative to the network system. Independent producers generally specialize in making original "live" programs, or "film" programs especially for television; or they distribute films originally made for theatrical exhibition. "Film syndicators" sell both types of film programs to advertisers, stations and networks, in series scheduled for 13, 26, 39 or 52 week periods, broadcast weekly at the same time, and consisting of integrated half-hour segments, including time for commercials. Producers and syndicators sometimes sell programs to the networks outright, or they sell proprietary interests in programs which the network in turn sells to network advertisers with a sure place in the network's schedule. But most non-network programs are sold to national station representatives and through them to spot advertisers. These independently produced programs and spot accounts provide the major source of programming and revenues for thirty-five independent stations, and for network affiliates in time periods which lack sponsors but where the network may not have agreed to provide sustaining shows, or where the station has rejected a network program and must either produce its own or turn elsewhere.

Here there emerge two important areas for inquiry — that of competition among the three national networks in rates and programming, and that of competition between them and the independent producer-spot advertiser system for choice station clearances. Regarding competition *among* the networks, on one hand, the main problem arises from barriers to new network entry, such as high interconnection costs, economies of large scale programming, network ownership of stations, financial incentives in the network-affiliate relationship and technical scarcities in the major markets. New network entry, considered only briefly in Section III, and largely ignored at the Hearings for reasons given below, would act to reduce the dependence of stations on any single affiliation and thereby perhaps to enhance program diversity.

On the other hand, regarding the networks' competition *with other components of the industry*, the problem to which we devote almost all of Section III arises from the special advantages that the networks gain in such competition from bargaining power they derive from their multiple station ownership, their role as station representative for their own stations and for certain affiliates too, and their option privileges, must-buy, affiliation and disaffiliation criteria, rate policies, sliding-scale compensation schemes, etc. Assuming that technical barriers to entry are relatively insoluble in the near future, the alternatives are either to promote more competition between the

networks and the national spot-independent producer system, or to turn to other institutional arrangements like Commission regulation of rates, subscription television, government licensing or ownership of the networks, federal subsidies, etc.

A final characteristic of the television industry is its legal status: it is "affected with a public interest." Stations, though not networks, operate under licenses granted by the Federal Communications Commission, for three-year periods, to applicants whose legal, technical and financial qualifications, program plans, past experience, and involvement in the community, appear to qualify them to serve the public interest. Under the Communications Act they are expected to serve as informational and instructional media as well as disseminators of entertainment and advertising. Although networks are not similarly licensed the Commission affects their operations by what it requires station affiliates to do and not to do in executing their public service responsibilities. One major aspect of the public interest with which the FCC is concerned is that of "program diversity" — diversity in terms of cultural, social, political and economic viewpoints expressed therein. In implementing program diversity FCC has sought to:⁵ (a) diversify station ownership through ceilings on multiple station ownership and rules on cross channel mergers; (b) fortify licensee independence by rules governing affiliation contracts and criteria to be employed in renewing broadcast licenses; (c) encourage the entry of local residents as opposed to outsiders unacquainted with community needs, the better to promote programming of local interest; (d) maintain competition among different components of the industry. But the Commission also weighs heavily in its licensing and allocation decisions past broadcast experience and adequate program resources. Hence there are sometimes conflicts between policies to strengthen the structural conditions (competition, ownership diversification) that facilitate program diversity, and those that promote the efficiency, stability and growth of the station and network resources needed for the desired programming.

In this context we may restate our purpose here as follows: the examination of issues at stake in current proposals to alter the TV channel allocations (see Section II below) and the present network rules (Section III below), proposals which focus largely on the maintenance of competition but which really bear on the Commission's other

5. See H. S. Levy, "Diversification and the Public Interest," *Yale Law Journal*, Jan. 1957, pp. 365-96; also my "Social Aspects of FCC's Broadcast Licensing Standards," in H. J. Levin (ed.), *Business Organization and Public Policy* (New York, Rinehart and Co., 1958), pp. 480-94.

proximate goals as well, and, of course, on the ultimate goal of program diversity.

II. UNDERUTILIZATION OF THE ULTRA HIGH FREQUENCIES

Significance of the UHF Band

Under existing channel allocations, a comprehensive geographic coverage and structural factors conducive to an adequate performance like a multiplicity of communities with local stations and of areas with competing signals, adequate educational transmissions, and effective network competition, all require an active UHF band and would therefore be threatened by any serious economic distress there. This was frequently stated at the Hearings.⁶

In this regard, we know that of 1875 commercial outlets allocated, only 556 (29.6 per cent) are in the Very High Frequencies, and a full 1319 (70.4 per cent) in the UHF, with UHF cities outnumbering VHF cities about three to one. Indeed the ratio is actually four or five to one in thirty-three states. Even more important, perhaps, the only commercial outlet, or the second and third outlets, in many of those states with eight or fewer VHF cities, are UHF outlets. In educational television, too, more than two-thirds of the 252 channels reserved for educators are in that band.

It is also generally accepted that establishment of a nationwide, competitive television service depends upon the use of both television bands. Since the chances for three networks to sell time during the prime evening hours (7:30–10:30 P.M.) are greatest in markets with three or more stations, it is crucial that only 35 cities have three or more VHF stations, whereas if we consider UHF channels too, 57 of the top 100 markets have *four* or more channels and a full 80 markets have *three* or more. In short, technical factors alone militate against the success of more than two national networks unless UHF as well as VHF outlets survive, and this was widely acknowledged at the Hearings.

So far as *actual* as opposed to *potential* performance goes, moreover, a full 96 per cent of Americans now live within the range of at least one television signal and so geographic coverage has clearly been maximized. Yet only some 350 of the 1240 cities where channels are allocated actually have a local station; and only 63 per cent of the people live in markets allocated three or more outlets (even though 83 per cent live in markets economically able to support them). Hence the local station and multiple service objectives are still to be achieved.

6. See *Television Inquiry*, II, 757–59, 764–65; *UHF Hearings*, pp. 130–33, 377–78, 519–23, 939–44; also *Celler Report*, pp. 6–9.

Furthermore only three national networks remain now that DuMont has withdrawn; and of thirty-one educational TV stations operating, only seven are UHF outlets even though two-thirds of the reserved channels are in that band.

In explaining this pattern, it is generally agreed that inadequate economic support in small markets and in markets near large urban centers, where an overlapping in retail trade areas and TV signals is hard to avoid, has operated to limit the number of local community stations. But it is also held that such obstacles may in part subside as those markets grow in population and buying power. Beyond this, however, the serious distress of UHF stations is considered a major problem, the failure to resolve which might well jeopardize not only the local station goal but all other policy objectives as well.

UHF Troubles

That there is no doubt about the financial troubles of UHF stations is clear from frequent testimony of industry and government officials, and from Table I, below, which indicates that in April 1958, some 83 per cent of 510 commercial TV stations on the air operated on VHF channels, even though less than 30 per cent of all channels allocated were in that band. Stated another way, UHF television with more than twice as many outlets allocated as VHF, accounted for less than one-fifth of the total number of operating stations. This alone suggests the unattractiveness of UHF channels. In addition, of 177 permits to build TV stations that discouraged enterprises returned unused to the FCC between 1952 and 1958, 144 were for UHF stations and only 33 for VHF stations.⁷

TABLE I

TOTAL OPERATING TELEVISION STATIONS, APRIL 1958

	VHF	UHF	Total
Commercial	424	86	510
Noncommercial, educational	24	7	31

Source: Broadcasting Magazine.

The key factors underlying the UHF troubles can be characterized briefly as follows:⁸ Without adequate resources or affiliations with a major network, UHF stations cannot transmit programs of

7. Comparative profit and loss data further reveal UHF's losses in contrast with VHF — even holding constant factors like community size, network affiliation, station revenues, etc. (See *Television Inquiry*, I, 159-65.) Losses for 95 UHF stations actually fell from \$10 million in 1954 to \$1.9 million in 1956; but during this time profits for 269 "post-freeze" VHF stations rose to \$16.6 million.

8. See *UHF Hearings*, pp. 147-49, 203-9, 240, 244-45.

wide appeal and people will therefore not pay extra to convert their existing television sets to receive UHF signals or to buy all-channel receivers. The reluctance to convert is also accentuated by UHF's limited range. Without adequate set coverage UHF stations then find it hard to secure choice network and advertising affiliations, their programs will deteriorate still further for lack of adequate resources, and the vicious circle is complete.

These are the factors which explain the tendency, in Table II below, for average revenues per UHF station, the percentage of television homes with UHF reception and, to a lesser extent, the number of network hours carried per station, each to vary inversely with the number of VHF signals received by 50 per cent or more of television homes in the UHF market.⁹

TABLE II
UHF STATIONS CLASSIFIED BY NUMBER OF VHF STATIONS RECEIVED IN
UHF MARKET, 1954

Number of VHF stations received by 50 per cent or more of the TV homes in UHF market	Number of stations	Per cent of total homes with UHF sets (average per city)	Average monthly revenues (per station) ¹	Total network hours carried (per station) ²
0.....	12	35	\$32,000	23
1.....	32	25	21,000	17
2.....	12	18	20,000	14
3.....	23	16	13,000	16
Total.....	79	23	20,000	17

Source: FCC testimony at *UHF Hearings*, p.162, Table 4.

1. During period Jan.-Mar. 1954.

2. During week of Mar. 14-20, 1954.

By way of seeking a fuller explanation of the above trends (and of related data in footnote 9), it may help to review certain relevant points cited frequently during the Hearings. Five basic questions are really at issue.

1. Why hasn't the public converted more rapidly to all-channel receivers and why do new buyers not purchase such receivers at the outset?

*Factors cited:*¹ Consumer inertia. Failure of set manufacturers,

9. Other data reveal a high percentage of VHF stations reporting profits regardless of the number of VHF stations serving the area; whereas UHF stations reporting profits fall off very sharply in these VHF areas. (See FCC, Public Notice 23055, August 19, 1955, Tables 13 and 14.) We also know that in communities of over 250,000 population, fifteen VHF stations authorized after 1952 received about double the number of network program hours as did thirty-seven UHF stations, even though they had no great head-start, chronologically, in gaining the best network and advertising affiliations. (See *UHF Hearings*, p. 162.) As indicated, network programming is crucial to station success.

1. *UHF Hearings*, pp. 147-49, 240, 244-45, 530, 721-23, 741-42, 782, 1087-88.

heavily involved in VHF production, to publicize and promote all-channel receivers and converters. The high cost and technical defects of converters. Greater access to popular network programs on VHF stations. Superior VHF quality and signal strength under existing FCC rules. VHF's earlier start and the public's consequent investment in VHF-only sets.

2. Why do VHF stations get the bulk of network programming and advertising revenues and UHF stations get so little?

*Factors cited:*² To attract advertisers by giving them the greatest coverage for every dollar spent, networks prefer to affiliate with stations having extensive audiences. VHF stations are favored over UHF so far as VHF sets are generally much more numerous in most markets, VHF signal strength and clarity are greater, and VHF stations are older, more experienced and better established. Networks also prefer to deal with television stations owned by their old radio affiliates — and these are generally VHF outlets, too. Once the VHF stations gain affiliation contracts, the network practices guarantee steady advertising revenues.

3. How may FCC policies, particularly as put forth in its *Sixth Report*, have contributed to UHF's economic distress?

*Factors cited:*³ FCC's forty-four month cessation of new grants, needlessly long for the problems at hand, facilitated firm entrenchment of the first 108 VHF stations with the choicest network and advertising affiliations. FCC should not have placed VHF and UHF assignments in the same market areas (intermixture). FCC's concern with promoting rapid geographical coverage has come to conflict with its desire for local operations. The wider VHF coverage and higher power, necessary for rural service, have seriously threatened the solvency of local UHF stations even where no local VHF outlet actually operates.

4. What is the relative importance of the NBC, CBS, ABC networks and, formerly, of DuMont, in VHF and UHF television and how can it be explained?

*Factors cited:*⁴ In 1954 the following clearances by four national television networks in the top 100 markets, 6 P.M. to midnight, indicate their relative importance in each band: 23 of DuMont's 39 clearances were on UHF stations; 31 of ABC's 51; but only 20 of CBS's 76 and 20 of NBC's 82. Long-standing NBC and CBS radio affiliations

2. *UHF Hearings*, pp. 973-75, 992-1000, 1087-88; *Monopoly Problems in Television*, IV, 6154-56; *Television Inquiry*, IV, 1761-80, 1838, 1856-78.

3. *UHF Hearings*, pp. 1018-20; *Television Inquiry*, I, 280-304; II, 330-43.

4. *UHF Hearings*, pp. 251-55, 972-73, 985, 1002-4, 1018-20.

in these most populous markets clearly favored them in gaining VHF clearance over Dumont — a newcomer to broadcasting and mainly a set manufacturer — and over ABC, struggling with inadequate resources since severance from NBC in 1943. Intangibles like customer loyalty, established contacts, knowledge of station needs, would tend to strengthen NBC's and CBS's ability to affiliate with the choicest VHF outlets (to say nothing of their greater resources, radio network experience and connections with advertisers). Once entrenched with VHF clearances, their ability to sell time to advertisers was further strengthened, thereby enhancing their program resources again, and so on. Never entirely resolved, however, is the question of whether DuMont's inability to gain more VHF clearances in the top 100 markets resulted more from the greater assiduity, ingenuity, radio experience and program superiority of her competitors, or from their restrictive tactics.

5. What is the role of the different importance of these networks in the UHF band in explaining its distress?

*Factors cited:*⁵ Ample NBC-CBS programs, and the prosperity of these networks, both reflect and help explain VHF's prosperity. Absence of such programs, and the availability of ABC and DuMont programs, similarly reflect and explain UHF's troubles. This would follow from the role these different networks play in servicing stations in the two bands. Research shows that the number of stations reporting losses, January–October 1954, was far greater when the station carried no NBC-CBS programming (or less than 7 and one-half hours a week), than when it carried more than 7 and one-half hours. This was true for VHF stations but was more pronounced for UHF stations. These trends are even clearer when UHF and VHF stations are classified according to average monthly revenues.

Policy

So much for the nature of the UHF problem, several of its implications, and the factors which have produced it. Remedies seek either to promote technical equality of the two bands (by equalizing power, antenna heights, signal clarity, etc.), to make markets homogeneous (by separating VHF and UHF markets, etc.), or to alter the network rules.⁶ The first two categories, backed in differing degrees by almost all parties at the Hearings, seek to equalize competitive

5. *UHF Hearings*, pp. 251–55, 803–9; *Television Inquiry*, I, 159–65, especially Tables 2, 5, 6, 15.

6. A fourth type of remedy includes institutional innovations like an outright federal subsidy for UHF stations, or a subscription system.

opportunities in television either by equalizing engineering conditions in both bands, or by avoiding the existing engineering inequalities. The objective is to make superior programming the major means whereby stations and networks compete for advertisers and audiences.

On the other hand, sharp disagreements arose over the tightening of network rules — the third category. The parties who would regulate affiliation criteria and the “right of first call” and who propose to curb exclusive-dealing, option time, and the must-buy list, sometimes imply that such practices, and not technical inequalities, constitute the main threat to full utilization of the UHF band and to the promotion of local stations and multiple service markets, and would endanger these objectives even if technical problems were resolved. They attribute at least part of television’s present structure — and its implementation — to alleged network discrimination against stations in small markets, against UHF stations in general, and against stations in markets overshadowed by powerful outlets in nearby cities.⁷

The major networks contend more successfully, however, that faulty channel allocations, and not restrictive practices, are the main culprit and that, at best, the FCC’s regulatory objectives are not fully consistent with each other.⁸ In other words, the networks attribute their practices to the underlying market structure, and the structure to factors like inadequate market support, uneven topography, FCC allocations and rules on signal quality, power, antenna heights, etc., in mixed markets, all of which provide a context within which the network’s legitimate attempt to maximize unduplicated circulation causes much of the trouble. This by no means implies that the network practices emerge unscathed as factors operating to restrain spot advertisers, affiliated and unaffiliated stations, non-network producers, etc. (See Section III below.) The point here is simply that their impact on the distress of UHF stations, the establishment of local stations, and on the growth of multiple service areas, is neither fully explored nor adequately established or refuted by the interested parties.

Turning next from the network rules to the proposals to equalize

7. *UHF Hearings*, pp. 153–54, 255–58; *Television Inquiry*, II, 762–65, IV, 1588–90.

8. See *Television Inquiry*, II, 759, 789–801, 814–26; IV, 2429–31, 2466–68; *Monopoly Problems in Television*, III, 4697–4701. ABC supported similar changes in allocations but also argued that, where new allocations were not forthcoming, FCC should curb “exclusive dealing” between key VHF stations and the two major networks along lines originally proposed by DuMont. (See *Television Inquiry*, II, 756–65, 777–78; IV, 2498.) For a refutation of the charges on network discrimination against UHF stations, overshadowed stations, small markets, etc., see CBS testimony in *Television Inquiry*, IV, 1780–84, 2047–71, 2155–56.

technical conditions and to make markets more homogeneous, alternative remedies were commendably examined in terms of the FCC's various policy objectives. Six points deserve brief mention.

First, tax relief on all-channel receivers received wide support as a way to reduce existing cost differentials between UHF and VHF sets. To the UHF operators, of course, tax relief was just another kind of badly needed assistance. But to the large networks and VHF representatives it was also a chance to break the vicious circle of economic distress without recourse to more elaborate, disruptive remedies that might involve detailed governmental regulations.⁹

Second, it was urged that VHF signals be limited to the station's retail trading area so as to prevent the overshadowing of UHF stations in nearby communities. This was supposed to favor the success of local stations — though a serious clash arose on whether it could be achieved without serious adverse effects on total geographic coverage and on the number of multiple service markets.¹ The consistency of all three objectives actually depends on the degree to which UHF stations are allowed to use boosters, at least until UHF power and antenna heights can be raised somewhat, and the degree to which VHF stations will be glad to relocate directly in the communities in which they are assigned and to use directional antennas to prevent overlapping. Once the islands of UHF stations begin to thrive, of course, tax relief on all-channel sets might further induce consumers to convert.

Third, selective separation of VHF and UHF markets was approved by most parties as a way to break through the vicious circle of distress.² The disagreements came over whether a minimal effort here would really suffice and whether a *full-scale* effort that separated the two kinds of markets widely might not seriously reduce service in fringe areas and also cost the public much for new sets and antennas.³

9. See *UHF Hearings*, pp. 191, 481–82, 770, 927, 980–81, 1042–43, 1076; *Television Inquiry*, II, 318, 333, 354. One of the strongest statements was by NBC (*ibid.*, II, 845–53).

1. In strong support of the proposal, see statement by Hometown Television Association, in *Television Inquiry*, II, especially 456–68; also *UHF Hearings*, pp. 190–91. An incisive rebuttal by CBS appears in *Television Inquiry*, II, 819–23.

2. See *Television Inquiry*, II, 313–16, 772–77, 853–55; *Monopoly Problems in Television*, I, 3270–71; *UHF Hearings*, pp. 977–80.

3. CBS estimated the cost of achieving the community station and multiple service goals, by giving *each* of the 1240 television communities either all VHF or all UHF outlets, with no overshadowing, as a billion dollars for set, equipment, and antenna conversion, plus the loss of service to several million viewers. Even short of that, DuMont's plan to give four homogeneous outlets to as many of 326 top markets as possible, and so to promote the multiple service objective, was

Fourth, the reduction of space between existing VHF stations was proposed to allow the construction of additional VHF outlets, the better to promote the local station and multiple service objectives — notwithstanding the continued competitive disadvantages of UHF stations. Yet because this solution might interfere with existing VHF signals, it was opposed as hurting the maximum coverage objective. One modified version sought to promote the multiple service objective, with little or no harm to the other two goals, by proposing selected changes in the location of VHF stations, modifications of power antenna heights, the use of directional antennas and other technical changes, thereby raising the number of major markets with three or more comparable markets from 52 to 84.⁴ The main objection was that it would do little to bring the UHF band into full utilization even though, by helping three networks to prosper, it might promote the multiple service objective.⁵

The *fifth* and most radical proposal was to move all television to the UHF band. Only this could supposedly promote the FCC's major regulatory objectives simultaneously, fully activate the whole UHF band, and release twelve scarce VHF channels for other uses. To be sure there were fears that up to 8 million people would lose service in fringe areas and that the transition would cost from \$1.5 to \$2.5 billions. But the use of boosters, satellites and higher powered UHF assignments was proposed to minimize the loss of service; and a ruling that stations must transmit UHF-VHF signals simultaneously during the conversion period might help reduce the public's dollar costs.⁶

found to help the local station objective only a little (due to overshadowing), to deprive a million people of service, to cost the public \$200 million for set conversion and antennas, and another \$30 million for new station equipment. (See *Television Inquiry*, II, 797-808.)

4. Another version noted that three more VHF channels would provide three or more competitive outlets to *each* of the top 100 markets; whereas seven more VHF would provide four or more outlets, thereby facilitating the survival of four networks. Of course, some set conversion would be needed, but no new antennas. The main objection was that UHF would again be left idle, that VHF channels were badly needed for governmental and military use, and that any inroads on the VHF channels reserved for educators would be highly undesirable.

5. For the different versions of this proposal, see CBS testimony in *Television Inquiry*, II, 810-19, 827-28. For criticism of the CBS plan, see statements by the FCC in *Monopoly Problems in Television*, I, 3269-70, by NBC in *Television Inquiry*, II, 859-61 and by Hometown Television Association, *ibid.*, II, 452-54.

6. The case for moving all television to the UHF band is set forth in *Television Inquiry*, I, 291-94. The most systematic estimate of costs appears in the CBS exhibit in *Monopoly Problems in Television*, III, 5036-44, and in its statement in *UHF Hearings*, 975-76. Lastly see FCC's balanced commentary on the data needed to plan such a move and to assess its net advantages in *Monopoly Problems in Television*, I, 3269-74.

A curious clash came over the *sixth* and final proposal, viz., that existing ceilings on multiple station ownership be waived, to some extent, wherever a network or non-network organization sought to purchase or build additional UHF outlets. With his abundant resources, the multiple owner would supposedly be able to survive the long period of losses necessary to put UHF on its feet. This was also a sure way to get essential network programming and advertising revenues for UHF stations — for the station would either be owned outright by a network, or its parent company might at least have the bargaining power to secure network affiliation and perhaps even to get it on the must-buy list.⁷

On the other hand, the proposal was widely opposed for fear that single station owners would generally lose to the multiple owners in contests for advertising support, for choice network affiliations and special agreements on rates, compensation and must-buy status, and for discounts on commissions paid to spot representatives and in prices paid to film syndicators.⁸ Multiple station ownership was also feared to jeopardize the growth of locally owned and operated stations and thereby to limit the extent of local programming.⁹

The upshot of these lengthy deliberations was threefold. *First*, the FCC was left free to continue its long-run exploration of the possible movement of all television to the UHF band. The interim policy was to be one of selected separation of UHF and VHF markets, and case-by-case decisions on VHF-UHF power, antenna heights, the use of booster stations and directional antennas, on relocations and reductions of mileage separations to facilitate the introduction of new VHF outlets. But there has been no subsequent attempt to reduce taxes on all-channel receivers, or to earmark more VHF channels for commercial TV. *Second*, we are strongly cautioned against any further relaxation of multiple ownership rules for fear of adverse effects on competition and local ownership. The possibility that

7. A strong case for relaxing the rules is made in *UHF Hearings*, pp. 257, 736-38, 742, 927, 947-48, 981, 1044, and in *Monopoly Problems in Television*, IV, 6111-12, 6117-18.

8. See *Celler Report*, pp. 34-39, 84, 141; also important memo from the FCC's Network Study Committee, in *Monopoly Problems in Television*, I, 3761-84, especially 3771-73.

9. Nor would the countervailing power that multiple station owners exercise against the networks necessarily guarantee any compensating improvement in performance (see *Monopoly Problems in Television*, I, 3773-75). For big non-network multiple owners sometimes suffer in transactions with the networks notwithstanding their alleged market power; and transfers exacted in cases of bilateral oligopoly imply no necessary social gain anyway. (See *Celler Report*, pp. 97-108 and details on the NBC-Westinghouse exchange in *Monopoly Problems in Television*, I, 3155-3205.)

looser rules might act to bolster the flow of venture capital into UHF and eventually make stations in that band comparable with those in the VHF band, thereby facilitating new network entry, was found to be more than counterbalanced by the adverse consequences of multiple ownership on other components of the industry. *Third*, and most important, the possibility of modifying the network rules was to be explored in its own right, *wholly aside from the UHF problem and its possible remedies*. Indeed, this brings us to our next section.

III. NETWORK PRACTICES AND PUBLIC POLICY

Notwithstanding DuMont's spectacular exit in September 1955, the Hearings generally tend to ignore the impact of network practices on competition between the networks in rates and programming and their impact on new network entry.¹ Rather are the practices analyzed, for the most, on the assumption that — *even if allocations and the UHF problem remain unchanged, market growth is negligible, and no new networks enter* — new network rules may still help open up the prime evening hours to more independent producers, station representatives, spot advertisers, and independent stations, in a way that enhances the financial support and program resources available for independent stations, as well as the ability and willingness of network affiliates to carry more local programming, and that facilitates competition between the industry's several components.

In other words, the objective is to secure an improved performance from the *existing* technical-economic structure — wholly aside from the possibilities of altering that structure or of supplementing it with institutional variations like subscription services. Nor does the network claim, that new allocations and a growing market will gradually resolve many "abuses" now attributed to their practices, necessarily challenge the wisdom of an interim policy. Granted, of course, that such a policy's *long-run advisability* would have been clarified by careful study of the thesis that network practices had contributed to, as well as resulted from, the UHF distress, the growth of multiple station ownership, the blockading of new network entry, etc., and

1. To be sure, there are brief comments on the latter, like those in *Monopoly Problems in Television*, I, 3154–55, 3726, and especially those by DuMont in *ibid.*, II, 4491–4520. But note also the absence of any thorough statement at other points where one might have been expected, such as *ibid.*, I, 3207–13, in *Celler Report*, *passim*., and *Magunson Report*, *passim*. Even the FCC's recent *Network Study*, which explored many factors affecting new network entry very carefully (see its chap. IV), rather ignored the implications of market growth and took the existing allocation of frequencies as a constant. (See this writer's review article, "Workable Competition and Regulatory Policy in Television," *Land Economics*, May 1958.)

would restrain access to the prime markets regardless of future changes in allocations and market support.

Much of the elaborate testimony on competition between the networks and other components of the industry really bears on three fundamental questions which we examine in the rest of this article. Do the network practices — especially option time and the must-buy rule — act to bar advertisers, independent stations or independent producers from any important television market, or to restrain the affiliates' freedom of choice? If so, do they result in a compensatory improvement in television's performance, such that justifies them nonetheless? If not, would the proposed remedies so disrupt network operations as to harm and not to promote the public interest?²

A. Are the Network Practices Restrictive?

The main evidence examined on this first question was the impressive success of networks in clearing programs on their affiliates;³ the rise in profit-participation arrangements between networks and independent producers where the former gain interests ranging from 25 to 50 per cent; the growth of outright program sales;⁴ the decline in the number of TV film producers and in TV film output;⁵ and profit analysis for 1955 revealing rates of return on net assets of 44 per cent for CBS and 35 per cent for NBC, and rates on tangible property of 99 per cent and 88 per cent respectively.⁶

Independent station owners and program suppliers attributed such trends to the competitive advantages that the networks' practices gave them over the independent producer-spot advertiser system, in reaching the choicest markets at the best times, and to their role in blockading new network entry. Needless to say the networks sought to refute these charges.

The attack on network practices was fivefold. First, option

2. One final query, largely ignored except for the interesting testimony on subscription television, is whether such alternative institutional arrangements are better ways to enhance the industry's performance than new chain rules or new allocations applied to the present system of advertiser-supported networking.

3. One study of 40 cities in which NBC and CBS had 80 basic affiliates, showed that 93.7 per cent of all half-hour programs ordered in the three evening option hours were cleared. A second study of the 52 stations on CBS's must-buy list reported that 8 stations cleared all 41 of the half-hour programs ordered during February 1957, that 13 cleared all but one, and that 40 stations cleared at least 36 of them. (See *Television Inquiry*, IV, 1526, 2920-21, and *Magnuson Report*, pp. 26-27.)

4. See *Celler Report*, pp. 40-50; *Magnuson Report*, pp. 73-75; also *Monopoly Problems in Television*, II 4021-48; IV, 5665-73.

5. See *Magnuson Report*, pp. 40-41.

6. See *Celler Report*, pp. 24-34.

time was said to explain the networks' great success in gaining prime clearances in so far as it forces affiliates to take at least some unwanted programs, reduces their incentives to innovate local programming — either commercial or sustaining — and may displace non-network advertisers and producers without notice.⁷ Second, the mererisk that networks can invoke option privileges to displace them from *unused* option time was held to have had turned advertisers from dealing directly with the affiliate's spot representative or with the independent producer, to dealing with the network, even though the latter requires them to order a basic group of stations including some in markets not wanted.⁸ Third, it was claimed independent producers often agreed to sell the network interests in their programs to avoid such displacement risks and the probability that networks will favor their own programs in borderline cases, with a further curtailment of independent program production resulting.⁹

Fourth, the network practices were said to put independent stations at serious competitive disadvantages on several counts. Because such stations cannot afford to produce all the programming they need to build audiences, to spread overhead, or to attract advertisers, they must rely on outside programs produced by network or non-network organizations. But the network affiliate's first option on all network shows (i.e., the right of first call), prevents the unaffiliated station in the same market from carrying them — even though the affiliate clears them only on a delayed basis, at an off-hour, and though the independent station would clear them immediately, at a good hour, backed by a non-network sponsor. Moreover, the independent producer prefers to deal directly with the networks, rather than with the independent station, not only to avoid the displacement risks mentioned above, but for fear that he will otherwise jeopardize his clearances in markets which lack independent stations. Lastly, the national-regional advertiser who might prefer to sponsor an available non-network show, on an independent station anxious to handle it, may be blocked from doing so by the must-buy practice which requires him to place his order elsewhere.¹ Fifth, the must-buy

7. *Television Inquiry*, IV, 1477, 1483, 1486-89, 1505-7.

8. *Television Inquiry*, IV, 1490, 1494-95; *Magnuson Report*, pp. 46-47.

9. On the networks' power to exact profit-sharing arrangements, see *Monopoly Problems in Television*, II, 3980-93, 4020-24; 4068-69. On the networks' tendency to favor their own programs, see *ibid.*, II, 4000-1, 4066-68; IV, 5703-11, 5722-24, 6068-69. Network affiliates were also accused of favoring network to non-network programs, other things equal, and again the networks challenged the charge. (*Monopoly Problems in Television*, II, 4032-36; IV, 6055, 6059; *Television Inquiry*, IV, 2670-75.)

1. *Television Inquiry*, IV, 1495-99, 1520, 1539-40.

practice was further said to bar certain advertisers from prime markets because they cannot afford to order all required stations, or because their market needs diverge widely from the list. Nor was unused option time, or time on independent stations, considered any adequate substitute.²

In short, to gain secure national-regional clearances, during the prime hours, independent producers must sell through those very networks that compete with them in program production, and not directly to affiliated or independent stations where time may be available, or to spot-local advertisers who may be willing to sponsor the programs in selected markets. In other words, the network practices act to prevent independent stations, producers and advertisers from dealing directly with each other to arrange national-regional distribution. The result is a restriction of the public's program alternatives and disproportionate network profits.

By way of rebuttal, on the other hand, the networks claimed that their superior, popular programming largely explained their success in gaining program clearances in both option and nonoption periods; that most affiliates would have granted the clearances even without the option privileges; that affiliates retained the ultimate power to reject network requests and had sometimes exercised it; that options had never been legally enforced to displace spot programs; and that the real reason for the rise in profit-sharing schemes was simply that the networks had invested heavily, and borne sizeable risks, in the launching of non-network shows and that such arrangements were logical outcomes.³ Furthermore, they held that non-network program sources were numerous and spot business was growing notwithstanding contentions to the contrary; that no network seeking maximum revenues could really afford to favor its own shows, over non-network shows, unless they clearly merited it; that the must-buy list was less restrictive than feared, what with advertisers often ordering even more than the required stations, and with few, if any, complaints ever heard.⁴

Regarding the network profit rates, finally, it was shown that these were not excessive, even if computed as percentages of net assets or of tangible property, *when compared with rates earned in other*

2. *Television Inquiry*, IV, 1493-95.

3. On the above points see *Television Inquiry*, IV, 2303, 2404, 2444-45, 2495-98, 2536-37, 2565-66, 2579, 2641. Note also that NBC reported that during the first quarter of 1956, 103 affiliates refused network requests for clearances at the rate of 3.9 programs per station.

4. On these matters see *Television Inquiry*, IV, 1790, 2174-75, 2181-82, 2236, 2418-19.

comparable industries. But more important, broadcasting was said to require heavy, risky investments in the creation and acquisition of programs, the organization of program structure, and in time sales to advertisers — in addition to investments in physical assets like studio equipment and transmission facilities. More appropriate for industries with such characteristics, the argument went, were operating ratios of expenses to revenues which, for CBS, were only 4.6 per cent after taxes, in 1954, and, on a cumulative basis through 1955, only .45 per cent — compared with an average of 6.1 per cent for 2400 leading nonfinancial corporations in 1954, and rates of 3.9–12.6 per cent approved for certain utilities.⁵

Briefly, then, the networks contended that restraints imposed by their practices were negligible, at most, and, as will be seen later, more than justified by the social benefits resulting.

In evaluating these conflicting arguments, the investigating committees and subsequent *Network Study* endorsed outright much of the critical attack,⁶ but also hastened to reassure the networks that no proposed rule change would harm their financial position seriously. The Magnuson Committee actually waivers between a theory of outright coercion, for which there is little explicit evidence in the testimony of affiliates and advertisers who seem basically satisfied, and a theory that the network-affiliate relationship is crucial.⁷ In the latter case, the network's power and the profitability of sharing its strategic position would explain the frequency of its prime clearances, the difficulties of spot advertisers and independent producers in entering prime markets, etc. The question really at issue here as elsewhere in the Committee reports is whether deliberate coercion, or pressures inherent in the network-affiliate relationship, or the affiliate's imperfect knowledge of network programming, or some combination of all three factors, best explains the high rate of network clearances and the resulting adverse effects on other industry components.

Regarding adequate market information, better knowledge of *all* program alternatives might conceivably tip the balance in borderline cases where affiliates and advertisers are now indifferent as to which of two programs they select. But the Magnuson Committee's frequent appeals that affiliates evaluate network and non-network programs more carefully before choosing, sound rather naive unless one gives considerable weight to the "personal element" as a determi-

5. *Monopoly Problems in Television*, III, 5419–29.

6. See *Magnuson Report*, pp. 24–25, 28, 44–46, 49–51, 56–57, 73–77, 97–101; *Celler Report*, pp. 40–45, 78–94.

7. See *Magnuson Report*, pp. 18–22, 37–38, 66–67.

nant of economic behavior in market structures like this. The real danger, of course, is that where program alternatives are roughly equal — and it is often hard to distinguish between them — advertisers and stations would tend to prefer the network program on grounds of long-run business relations. Nor should we be surprised if, other things equal, networks favored programs from their own production departments.⁸

To test the theory of coercion we actually need, among other things, statistical analysis of program clearances by size of market and of station, in option and non-option time periods, and for stations with and without network affiliations, on and off the must-buy list.⁹ Furthermore we must consider whether the fewness of networks somehow exerts pressures that would exist *even if option time, must-buy and other practices were modified drastically*. For if the network-affiliate relationship, and no specific practice, really underlies the clearances, then elimination of such practices would not open up these prime hours for competition.

This point was not really handled satisfactorily at the Hearings. It raises other questions: How efficient are the three networks? How much room for additional networks is there, assuming each would reach an optimum size? Would additional networks and an increased number of comparable facilities in the top markets reduce general dependence of stations on network affiliation, or any single affiliation? As indicated earlier, this takes us to the whole question of why a fourth network disappeared and of whether one or more newcomers may enter in the foreseeable future. Lastly one wonders about the relative importance in network-affiliate relations, and in gaining choice clearances, of the network's power to manipulate station time rates and compensation arrangements, to withhold or withdraw affiliation contracts, to put stations on or off the basic required list, to invoke option privileges, or to order the networks' owned and operated stations to institute certain policies on rates, programming etc.¹

8. For evidence see *Magnuson Report*, pp. 67-73; also *Television Inquiry*, IV, 2670-75.

9. The FCC *Network Study* has in fact undertaken such analysis with the conclusion that the importance of these practices has been overrated by both sides, although they do have some limited effects. More important is their indirect, intangible role as talking points in bargains struck between the networks and their affiliates. (See the writer's review article in *Land Economics*, May 1958.)

1. The *Network Study* refers to such factors as props to network bargaining power and emphasizes that although no single practice has serious coercive effects, together they constitute an important lever, entrenching the networks against the competition of spot advertisers and independent program producers. In this

B. Is There a Compensating Improvement in Television's Performance?

The networks not only deny that their practices are seriously restrictive, but they go further and defend them as essential to networking and to the unique service that results. The upshot of the argument, although never stated so extremely, is that any tendency to exclude independent producers, smaller advertisers, and independent stations, is more than compensated by the high quality of network programming. Less sympathetically, one might say that the networks try to have it both ways, maintaining first that their practices are *not* restrictive, and then justifying them even if they *are*.²

The uniqueness and social value of network program service, are set forth in terms of its comprehensiveness, its over-all balance in news, commentary, special events, public and cultural affairs, its role in innovating new program techniques and forms, and its simultaneous interconnection of widely separated points in ways that facilitate wide social participation and cohesion and that build steady audiences and a reliable medium for national advertising.³ Nor were these vital attributes ever seriously discounted by the network critics, the investigating committees, or the subsequent *Network Study*, except indirectly so far as they also lauded the social importance of local, community programming. The main argument arose not over the value of these acknowledged accomplishments of networking (or over its crucial contribution to the industry's spectacular growth), but rather over the role of the network practices in implementing the above. The proposed rule changes were defended, in the face of deep network opposition, as eliminating certain restrictive aspects of their practices, without seriously upsetting their finances or their social functions.

Briefly the network argument is that the network service is costly and requires abundant, stable resources — resources that would be seriously eroded in the absence of the trade practices under attack. The practices, in turn, are said to arise from three of the industry's basic economic characteristics, viz., television's need to compete against other national advertising media like magazines, regard, the *Celler Report* notes the anticompetitive impact of large quantity discounts (unrelated to real cost savings) offered to big advertisers in all network time sales (*Celler Report*, pp. 61–68); and there were frequent proposals to regulate the networks' affiliation practices and rate policies (*Television Inquiry*, IV, 1516–17, 1560–65, 1837–40, 2410–14).

2. This really runs throughout the network testimony, but is clearest in their antitrust briefs in *Television Inquiry*, IV, especially 1958–94, 2368–88.

3. On the uniqueness of network service, see *Television Inquiry*, IV, 1713–16, 1958–62, 2162–63, 2294–98, 2393–97, 2485–86, 2499. On the case for live vs. film programming see *Television Inquiry*, IV, 1813–14, 2277–89.

newspapers and radio, where without some guarantee of simultaneous nationwide clearances advertisers might turn elsewhere; the special risks implicit in the network cost function, arising from the fact that should sponsors drop programs or fail to order a station, this not only reduces network and station revenues, but also raises the network's costs so far as it must provide sustaining programs to fill in and must continue to bear the high interconnection costs alone — such that the loss of a few big network advertisers on a few key shows may mean the difference between an annual profit or loss; the additional uncertainties that characterize creative industries, where long-run investments in talent, programming, equipment, are needed for innovation and experimentation, and where the networks alone stand to bear the losses from faulty investment decisions, though the gains are shared by the stations.⁴

The networks imply that their practices cushion them against such uncertainties, without operating to exact exorbitant profits. Nor can we really evaluate this contention without first considering the economics of networking. For instance, does option time, as is claimed, prevent those piecemeal defections of stations, particularly in a declining market, which would act to turn the network advertiser elsewhere for nationwide circulation and which, by forcing the network to broadcast more sustaining programs notwithstanding high interconnection costs, might finally lead to a curtailed service? Does it also forestall the possibility that stations which the networks drop and which must turn to spot or local advertisers and to non-network producers, might not rejoin the network when markets grow stronger? The must-buy list has been similarly defended as helping the network to safeguard its expensive investments in interconnection facilities. If such claims and related ones are valid, then together these practices, and others like profit-sharing agreements, might well help stabilize network revenues, spread high interconnection costs over more time sales, compensate for the great risks which creative industries like this must bear, and provide the abundant resources needed for the networks' vital social functions.

Unfortunately these matters were not explored more systematically — on their own ground. Rather did the rebuttal take a less direct line, if a cogent one; namely, that if affiliates and advertisers would prefer network to non-network programs anyway, on grounds

4. On the industry's peculiar risks, its cost function, and other special economic characteristics, see *Television Inquiry*, IV, 1753–54, 1759–62, 1778, 2301, 2424–28. On the reasonableness and essentiality of the network practices in this context, see *Television Inquiry*, IV, 1805–37, 2290–2316.

of profitability, suitability, popularity, and the increased value of adjacent time periods to spot advertisers, as frequently claimed, then the practices were surely superfluous and should not be maintained so tenaciously; that if the practices were in fact crucial to network stability, they were probably restrictive too; and that even the social value of network service did not justify an unregulated restrictionism.⁵

Indeed, regarding the first assertion, one cannot help but wonder, with the network critics, how the networks could possibly suffer serious losses if their program quality, and not coercive tactics, was really crucial in gaining program clearances and advertising revenues — unless, of course, borderline decisions between network and non-network shows are frequent and the practices “tipped the balance” without being “seriously restrictive.” Concerning the social value of network service, moreover, one must also agree with the investigators that where public policy endorses restrictionism it ought somehow to control the resulting profits and performance even though this is not always done. In this industry, judging from the range of proposals heard, control might well entail the direct licensing of networks, their subjection to the same periodic program reviews to which all licensed stations are now subject, and a regulation of affiliation criteria, rate policies, compensation schemes, and so on. Clearly the networks never really faced fully the implications of their contention that the social value of their service justified any incidental restrictionism. But one wonders whether the Committees or the *Network Study* completely established their own major thesis; viz., that practices which stabilize revenues most likely restrain trade in some undesirable sense too, *but that new rules can be formulated to reduce the restrictionism without seriously jeopardizing network finances or network program service.*

C. Are the Proposed Rule Changes Reasonable?

The line between trade practices which aim to provide a modicum of revenue stability in the face of unusual market uncertainties, and outright restrictionism to exact monopoly gain beyond that, is undoubtedly hard to draw. Yet many of the sharpest disagreements between government and business can be resolved only if such a line is drawn. Without a systematic analysis of how the network practices have operated in the face of economic fluctuations, and without a more thorough inquiry into the risks implicit in long-run talent contracts, the launching of stations and programs before sets were in wide use, the special network cost-function, and the high interconnec-

5. *Magnuson Report*, pp. 28, 31, 36–38, 45, 61.

tion costs, one wonders how the reasonableness of the proposed rule changes, no less than of the network practices in 'dispute, can be established convincingly. Should such studies support the investigators, on the other hand, this might well calm the network's fears more than does the mere repetition of verbal assurances, assuming, of course, that the networks do really seek no more than a reasonable economic stability.

As things now stand, the investigators do the next best thing to launching such studies. Within the context of existing technical and market conditions they would promote a competition that acts to diversify and enrich television programming because it is consistent with network financial stability. The tentative compromise reached may not please all interested parties; but it is a compromise nonetheless, and perhaps the most that can be expected at present.

Thus the Committees propose to eliminate must-buy and option time, and to formulate a rule that defines and prohibits "exclusive dealing" between stations and networks or other program suppliers. But they also accept as a substitute for must-buy, a minimum-buy provision whereby networks might require advertisers to spend a certain sum if they are to use the networks at all, but without permitting them to specify *which group* of stations must be ordered at the desired time period. This is expected to eliminate the more restrictive aspects of must-buy without unduly upsetting network finances.⁶

Indeed the *Network Study* compromises even more. Besides endorsing this action on minimum-buy, it also explicitly *rejects* the proposal on "exclusive dealing" to avoid undue interference with business operations. Regarding proposals to regulate the networks' affiliation and disaffiliation criteria, rate policies and compensation arrangements, moreover, the *Study* would simply require a *public divulgence* of the criteria and procedures the networks use, but no direct regulation, as yet. Furthermore, although it urges that networks be prohibited from serving as spot representatives for stations other than those they own and operate, and would restrict multiple station ownership in several ways, the *Study* would not prevent the networks outright from setting their affiliates' time rates, or from owning a full quota of five VHF and two UHF stations, including three VHF stations in the top twenty-five markets. Rather does it call for continued policing in these and related areas to catch abuses as they occur.

6. The committee argument is well-stated in *Magnuson Report*, pp. 36-37, 42-45, 100-1.

IV. CONCLUSION

Through their extensive inquiry into the UHF problem, multiple station ownership, and network practices, the investigators sought to formulate changes in the network rules and in channel allocations to implement not one but several policy objectives. Thus they sought to promote a freer entry of newcomers into television, an increase in local stations, and in competition of several sorts; the widest, most rapid growth in set ownership and geographic coverage of television service; the maximum number of multiple station communities and multiple service markets. They wished to pursue such goals, moreover, in ways consistent with a reasonable stability of network revenues.

Actually, however, the investigators treated these goals of efficiency, growth and stability, not as ends in themselves but as instrumental to a more diversified output. This diversity of output they saw as threatened by the UHF problem, by multiple station ownership, and by the network practices. They were particularly concerned over the lack of balance between local and national programs, between filmed and live matter, and over the deficient opportunities for independent program suppliers, advertisers and independent stations, to participate in the planning of program fare along with the national networks and big advertising agencies.

Now the major queries these Hearings raise are not unique to radio and television even though the form they take here is distinctive. Wherever entry is restricted due to technical and legal or economic factors, a more general question arises: Are the restraints on entry, and the degree of monopoly power they support, more than balanced by a compensating increase in the quality and quantity of output? Or would changes in the legal and institutional framework help improve the industry's performance? For instance, in air transport we start by asking whether restricted entry helps promote the industry's safety, efficiency, growth and contribution to national security, such as compensates for the increased market power it gives to the major trunk lines at the expense of the nonscheduled carriers. In motor transport the question is: do controlled entry and minimum rates stabilize revenues and act to improve the safety, regularity and general quality of service, such as counterbalances any tendency to lessen competition within the industry and vis-à-vis the railroads?

Here in radio and television the Committees are really trying to ascertain whether the restriction on entry needed to maintain satisfactory signal quality, the additional restriction imposed by technical scarcities and by the network practices arising therein, acted to

enhance not only the industry's profits but its program service too, in terms of geographic coverage, local-live programming, and general program diversity and balance. Their tentative conclusion, supported by the subsequent *Network Study*, was that it did not; that the blockaded entry was not fully justified by the resulting performance. Their proposed remedy, for the present, was to tighten the network rules as indicated, rather than to wait until a general economic expansion had reduced the magnitude of many of television's problems, *and rather than to institute drastic changes in the channel allocations or in the industry's present institutional framework based on advertiser support and national networking*. But this writer left the mountains of new testimony and valuable research feeling that although the rule changes were a wise interim measure, any long-run improvement in the industry's performance would require additional work in these other directions too.

PEAK LOADS AND EFFICIENT PRICING: COMMENT

By JACK HIRSHLEIFER*

In a recent article Professor Steiner presented what I believe to be essentially the correct solution of the problem of peak load efficient pricing.¹ The disagreement with Steiner's paper that calls forth this note does not turn so much upon the analysis, which I regard as a truly major contribution, but upon Steiner's *obiter dictum* to the effect that his solution involves discriminatory prices. The disagreement is, in some degree, a semantic one, turning upon the meaning of the word "discrimination." However, I shall contend that there is a fundamental underlying idea which permits us to tell when price differences are discriminatory and when they are not, and that under this concept Steiner's solution is nondiscriminatory. I shall, in addition, separate the short-run and long-run aspects of the problem and provide a more general solution (in this respect) than Steiner's — his analysis being limited to the long-run solution for a cost function of a particular shape. It is this more general result which makes apparent the essentially nondiscriminatory nature of Steiner's solution.

Our science advances in part as a result of refinement of thought which permits us to make such distinctions more consistently than they have been made in the past. In addition, there are certain important legal and policy implications in the use of the word "discrimination" that justify care on the part of the economic profession. In a nutshell I shall contend that the efficient price differences in a peak load situation shown in Steiner's analysis are not discriminatory because they are equal to the differences in the marginal cost of serving the classes of customers involved. By marginal cost I mean, ultimately, the marginal opportunity cost — the value set upon the resource by the customer in the most valuable alternative use being sacrificed. (It is necessary to bring in the more ultimate concept of opportunity costs because Steiner's solution is such as to leave marginal cost indeterminate in the usual, cash-outlay sense.)

The point of view adopted throughout this paper is that of an enterprise operated solely for "maximum welfare" rather than maximum profit. (We speak, of course, of "maximum welfare" or of

* University of Chicago.

1. P. O. Steiner, "Peak Loads and Efficient Pricing," this *Journal*, LXXI (Nov. 1957), 585-610.

“efficiency” in the old-fashioned welfare theory sense — that is to say, we accept the initial distribution of wealth and talent as a datum for our analysis, we assume that consumers’ preferences are the relevant criterion and that consumer decisions faithfully reflect these preferences, and so forth.)

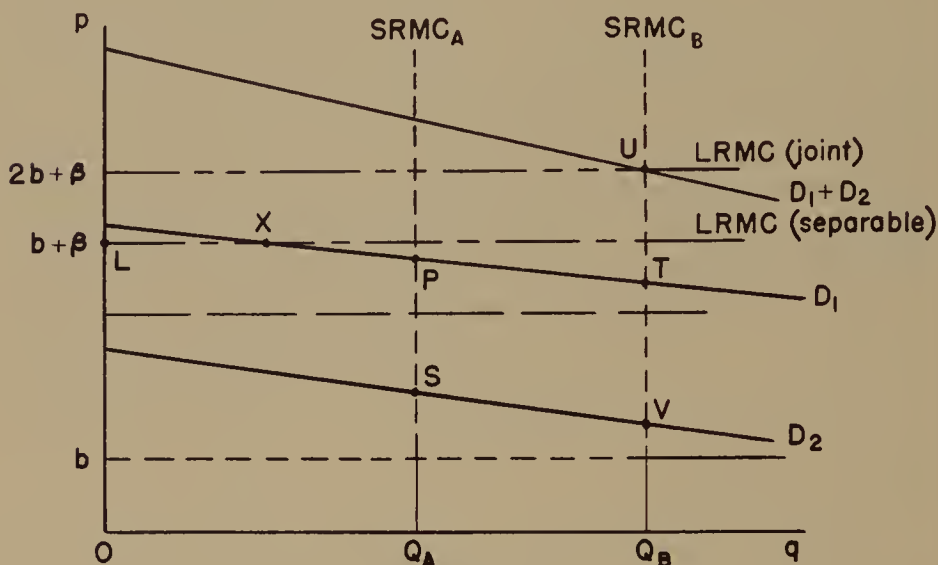


FIGURE I

SHORT-RUN AND LONG-RUN SOLUTIONS, DISCONTINUOUS COST FUNCTION

My interpretation of Steiner’s solution is illustrated in Figure I. These are two time periods (say, day and night) which generate the demand curves D_1 and D_2 ; these demands are assumed independent of one another. (This assumption is rather unreasonable, since off-peak and on-peak service may be close substitutes, but it is a valuable simplifying assumption for this problem.) The cost function is identical for each of the two periods considered separately: up to the fixed capacity limit existing, quantity q for either period can be provided at a cost of b per unit; beyond that point, demand for either of the periods can only be satisfied at a cost of $b + \beta$ per unit, where β is the cost per unit of providing new capacity. Demand for both periods combined can be satisfied at a cost of $2b$ per combined unit up to the capacity limit, and $2b + \beta$ per combined unit beyond that limit. In the simple analysis illustrated in Figure I, both b (variable cost per unit) and β (capacity cost per unit) are assumed constant. The presentation in Figure I differs from Steiner’s in that the latter’s demand curves are defined to be net of variable cost — that is, the amount b has been subtracted vertically from D_1 and D_2

as shown here, the resultant curves thus becoming "demand for capacity."² This definition is legitimate and highly convenient geometrically; the same simplification cannot be adopted here, however, as it obscures exactly that relation between the short-run and long-run solutions on which the present analysis centers. The demand curves used in this analysis are, then, ordinary demand curves for the commodity or service under consideration.

Previous analysts, outstanding among whom is Davidson,³ had tended to argue that the off-peak customers (D_2) should be charged only the variable cost per unit b ,⁴ while the on-peak customers (D_1) should be charged $b + \beta$ — variable cost plus capacity cost. The underlying idea is that it is the on-peak customer who is responsible for the purchase of the marginal unit of capacity. In the case illustrated, supposing that Q_A is the capacity in existence at a moment of time, such pricing would obviously be incorrect. Steiner says that "the peak shifts" (the on-peak customers at the price $b + \beta$ would take only LX , less than the off-peak customers); in addition, the off-peak customers would have to be rationed, since at the price b they would want far more than the capacity available. The correct short-run solution is to charge Q_AS and Q_AP to the slack and peak demand customers, respectively. If the price for either class of customer were higher, a quantity less than Q_A would be demanded. Since the demand curves are such that the values placed on an additional unit of q remain higher than the marginal cost b throughout the range of q from zero to Q_A , it is clear that the amount Q_A should be provided, and this amount would not be purchased if the prices were higher than Q_AS and Q_AP respectively. Prices lower than these would create excess demand, since more than the fixed capacity would be desired, and service would have to be rationed. As Steiner points out, Davidson was aware of the "peak-shifting" difficulty. He recommended certain empirical price adjustments where peak-shifting threatened, but Steiner is correct in charging him with vagueness on this score. In fact, the suggested adjustments are inconsistent with Davidson's principle of charging according to long-run marginal cost. (The correct principle, I shall argue shortly, involves charging according to the separable *short-run* marginal cost.)

2. Steiner, *op. cit.*, p. 588.

3. R. K. Davidson, *Price Discrimination in Selling Gas and Electricity* (Johns Hopkins, 1955), pp. 111–47. Despite my disagreement with Davidson on peak load pricing, I feel bound to say that the integration of theory with practical problems in his book represents a real high point in the application of economic theory to public utilities practice.

4. On the assumption that marginal cost equals average variable cost.

Steiner's solution was to point out that a welfare optimum requires the purchase of additional capacity until the vertical sum of the demand prices (on the curve labeled $D_1 + D_2$) equals $2b + \beta$ at Q_B , the prices Q_{BV} and Q_{BT} being charged. His *obiter dictum* was that these prices are discriminatory since they are based on demand differences, the cost functions being identical for the two classes of service.⁵ Actually, Steiner does not explicitly discuss the *short-run* problem whose solution was just given above (charging Q_{AS} and Q_{AP} for slack and peak demand), since he concentrates his attention upon the problem of purchasing and charging for "capacity." Had he separated this long-run problem from the short-run problem, he would have been led (I believe) to the correct short-run solution which is certainly implicit in his analysis.

Somewhat as an aside, I would like to mention that the usual engineer's or "Hopkinson" classification of costs into "commodity" costs and "capacity" costs (a third category is "customer" costs, which are not relevant here) has been rather uncritically accepted by economists. Such a classification only makes sense when it can be assumed, as Steiner does, that the "commodity" cost per unit is constant (b in Figure I), because in that case a charge equal to marginal commodity cost will also equal average commodity cost and so the charge will recover the total commodity cost. More generally, there will be a U-shaped curve, and a charge equal to marginal commodity cost may either exceed or fall short of average commodity cost. It will shortly be seen that in such a case efficient off-peak and on-peak pricing for the commodity or service makes it unnecessary to "charge for capacity" explicitly, in Steiner's terminology, or to "allocate capacity cost" in Davidson's language. That is, appropriate (short-run) marginal-cost pricing will be shown to lead not only to optimal short-run outputs but to the optimal level of capacity.⁶ It is true in practice, however, that it would be quite impossible to institute the continuously shifting commodity prices which would be necessary for the ideal solution as demand fluctuates from moment to moment — our assumption here of a single peak period and a single slack period is, of course, radically oversimplified. Given

5. "If the demand curves are different at \bar{x}_0 , the prices are unequal and since this is truly a case of joint costs, unequal prices in the face of equal outputs and joint costs mean discriminatory prices." Steiner, *op. cit.*, p. 590.

6. It is true, however, that under this arrangement total revenues may not cover total cost — so that, while the capacity is optimal under the proviso that the enterprise must stay in business, the question of whether or not it should be closed down entirely is left unanswered. This possible conflict between the marginal and the total conditions for an optimum is a familiar topic in welfare theory; the present paper is directed solely to the marginal conditions.

this assumption, however, correct short-run pricing makes capacity charges as such unnecessary.

My first difference with Steiner's argument, then, relates to his failure to separate the short-run and long-run aspects of the problem. The reason for this omission turns upon the cost function he employed. Thus his marginal cost function runs at the level b until capacity is reached and jumps to $b + \beta$ thereafter. But it is much more consistent with the nature of the production alternatives (and with traditional theory) to define in such a case a short-run marginal cost function $SRMC$ turning sharply upward from the level b at the existing capacity point (note the curves $SRMC_A$ and $SRMC_B$ I have drawn in Figure I), while the separable and joint long-run marginal cost curves are horizontal at the levels $b + \beta$ and $2b + \beta$, respectively.⁷ In practice one would expect something more like standard

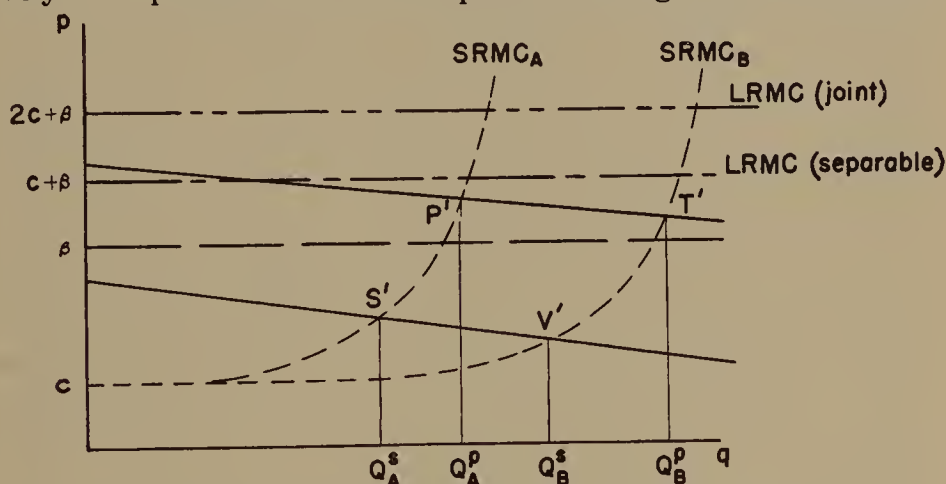


FIGURE II

SHORT-RUN AND LONG-RUN SOLUTIONS, TRADITIONAL (CONTINUOUS)
COST FUNCTION

curvature, illustrated in Figure II,⁸ but the discontinuous presentation employed by Steiner is a useful abstraction for a number of problems. The analogy between the accepted representation (Fig-

7. The joint long-run marginal cost curve refers to the cost per combined unit of production — a variable cost per unit of b is then incurred for each of the two periods in addition to the capacity cost per unit of β . The separable $LRMC$ curve indicates the cost of increasing output for one of the periods, output for the other period being held constant at some lower level. In this case only b per unit for one of the periods is incurred in addition to the capacity cost per unit of β .

8. Westfield's electricity study, for example, shows a U-shaped cost curve of traditional shape. See F. M. Westfield, "Marginal Analysis, Multi-Plant Firms, and Business Practice: An Example," this *Journal*, LXIX (May, 1955), 253-68.

ure II) and the idealized case Steiner considers (Figure I) should be noted, however. (Figure II does involve certain complications which will be explained later.)

Under the interpretation I would recommend, the following rules would achieve an equilibrium at a welfare optimum for the cost functions of both Figure I and Figure II: (1) Short-run: for both slack and peak demand, equate short-run marginal cost to the demand price.⁹ (2) Long-run: the rule here is somewhat complex because of the existence of two classes of cases. In addition, its application to the continuous cost function of Figure II must be deferred. The essential principle, however, is that new capacity should be purchased until the *relevant* long-run marginal cost equals the *relevant* vertical sum of the demand prices. (After establishment of the optimal capacity, prices should be set as in the short run — that is, the correct prices are the *SRMC* prices along the *SRMC* curve associated with the optimal capacity level.) The relevant *LRMC* curve will be “*LRMC* (joint)” in Steiner’s “shifting-peak” case — that is, when the demand curves are such that no single period’s demand alone justifies an increase of capacity. Diagrammatically, this means that (as in Figure I), at the existing capacity point the level of D for neither period alone exceeds $b + \beta$ — the separable long-run marginal cost. If the demand price D_1 for a single period alone did exceed $b + \beta$, an increase in capacity is obviously called for, and the relevant *LRMC* for comparison with D_1 is “*LRMC* (separable),” or $b + \beta$. Where this condition does not hold, an increase in capacity may still be indicated if the demand price for a combined increase of output ($D_1 + D_2$) exceeds “*LRMC* (joint)” — $2b + \beta$.

Despite the awkwardness of the verbal exposition, the fundamental idea is fairly simple and can be easily generalized to more

9. In the discontinuous case, the intersection of *SRMC* with the slack demand may, and with the peak demand must, occur in the vertical branch of the curve. In Figure I, both intersections are in the vertical branch.

There are two types of short-run solutions possible: the case illustrated in Figure I (for given capacity of Q_A) has both classes of customers taking the full capacity, while the other possibility would be a D_2 so low as to fall below the level of b before Q_A is reached — this latter is the less interesting “firm-peak” case. In the firm-peak case, the efficient price would be b for the slack demand customers. For the peak demand customers, Steiner agrees with Davidson that the price $b + \beta$ would be correct. This is true only in long-run equilibrium, however. In the short run, charging $b + \beta$ would not be generally correct. Looking at D_1 in Figure I with capacity Q_A in existence, the price $b + \beta$ would lead to only LX being taken, an obviously inefficient result. If in fact D_2 were so low as to fall below the level of b to the left of Q_A , it would be true that the capacity in place is excessive and in the long run should be cut back. But in the short run the efficient solution is to charge the price QAP , since the marginal value of service is greater than the marginal cost up to Q_A .

than two periods. Borrowing Steiner's language, we may say that more capacity should be provided if justified by the demand for any single period alone (i.e., if any D exceeds $b + \beta$) or justified by the combined demands of a number of periods (i.e., if $\sum_{i=1}^n D_i$ exceeds $nb + \beta$, where the n periods in question are taken to be those for which the level of D exceeds b at the existing capacity point).

The above long-run solution is simply an adaptation of Steiner's — it is the interpretation of the solution in terms of long-run and short-run marginal costs which represents the only difference here. Thus, I would describe Steiner's "firm-peak" case as the case for which "*LRMC*(separable)" is relevant; since demand off-peak is so low that nothing more would be produced off-peak even with an increase of capacity, the appropriate comparison is between the level of the peak period's demand curve and the long-run cost of producing more on-peak — $b + \beta$. For Steiner's "shifting-peak" case, "*LRMC* (joint)" is relevant; since more will be produced in both periods with an increase of capacity, the appropriate comparison is between $D_1 + D_2$ and $2b + \beta$. In brief, my contention is that properly looked at, outputs and prices in both short-run and long-run equilibrium are based upon equating the relevant demand prices to the relevant marginal costs — and hence are not discriminatory.

In a footnote in his paper, Steiner states his disagreement with this interpretation.¹ Rather than defend my position directly, which would involve showing that the discontinuous cost function under discussion is the limiting case of a continuous function as we increasingly kink it at the capacity point (compare Figures I and II) I would rather turn more fundamentally to the essential meaning of discrimination. In the same footnote Steiner mentions Professor Chamberlin's objection that the on-peak and off-peak services represent different commodities, so that the concept of discrimination is inappropriate. Steiner comments, however, that such an interpretation "virtually annihilates the concept of discrimination." I shall try to show that the concept of discrimination can usefully be defined for such situations, and that under this definition Steiner's solution is nondiscriminatory.

The following idea, I maintain, underlies the discrimination concept used in economic theory: discrimination between two classes of customers exists when the price differences charged are not equal to the differences between the cost of providing the marginal unit

1. Steiner, *op. cit.*, fn. 9, p. 590.

to customers in each class,² where by "cost" we ultimately mean the most valuable alternative foregone. In the usual textbook example, a price difference exists between two parts of an artificially divided market, there being no cost differences in supplying one market or the other. In that case the price difference is discriminatory. Suppose, however, that there *is* a cost difference between the two segments of the market. For example, let the commodity be water, supplied to one class of customer by gravity service, but to the other by pumping. In this case price equality would be discriminatory; nondiscriminatory prices would involve charging the latter category for the marginal pumping cost, in addition to the marginal commodity cost for the water. So mere price differences do not necessarily constitute discrimination, nor is price equality necessarily nondiscriminatory.

Turning back to the short-run solution at the capacity point Q_A , we see that there is a price difference: Q_AP is greater than Q_AS . It is easy to slip into an error here, however, and to say that there is no cost difference between serving the two demands, since the cost function is the same on-peak and off-peak. It is true that the cost *function* for serving the two classes is the same, but the effective point on the cost function differs. The discontinuity in the cost function as idealized in Figure I makes it impossible to define the short-run marginal cost in the usual cash-outlay sense, so we are forced back to the more ultimate opportunity-cost formulation. What is the opportunity cost of serving the marginal unit to some member of the class of customers represented by D_1 , when an aggregate amount Q_A is being supplied to the class as a whole? The opportunity cost is the most valuable alternative foregone, which in this case is the failure to serve some other customer in the same class willing to pay infinitesimally less than Q_AP for a unit of service. Similarly, for the class represented by D_2 , the marginal opportunity cost is Q_AS , or rather this amount less an infinitesimal epsilon.³ On this interpretation, the short-run solution arrived at involves price differences equal to the differences in the marginal opportunity cost

2. Some definitions of discrimination are based upon disproportions between prices and marginal costs in the separated markets, as opposed to the concept used here of an inequality between price differences and marginal cost differences. For the problem under discussion in this paper the distinction is inoperative since, it is contended here, the efficient prices are actually equal in each case to the relevant marginal cost. Consequently, the prices must be proportionate to the marginal costs, and discrimination does not exist under either definition.

3. This interpretation of marginal cost has been employed in W. Vickrey, "Some Objections to Marginal-Cost Pricing," *Journal of Political Economy*, LVI (June, 1948), 218-38 (at 231).

of serving the two classes, and thus can properly be said to be non-discriminatory. The short-run solution at Q_B , which happens to be the equilibrium or long-run solution, can be similarly analyzed.

At this point the reader may perhaps suspect that he has been the victim of legerdemain. Couldn't the same interpretation be used to defeat the standard textbook example of price discrimination? The answer is no! In the textbook example, the market is divided artificially. The commodity being the same, at the profit-maximizing solution the marginal customers in each class, while paying different prices, are being served at the same real marginal opportunity cost — the value of the first unit of unsatisfied demand in the *higher-priced* market is the most valuable alternative foregone. Therefore, no price difference is justified on opportunity cost grounds. In the case under consideration, the market division is not artificial — taking a unit away off-peak does not make it possible to supply a unit on-peak, so the higher on-peak value is not the relevant alternative social opportunity for the off-peak service. In this sense, Chamberlin's objection that off-peak and on-peak service are not the same commodity pinpointed the source of error in Steiner's conclusion that his price differences are discriminatory.

Actually, the abstraction of the discontinuous cost function is what is causing most of the difficulty here. If we present essentially the same picture with a continuous function, as in Figure II, the standard short-run solution is to set the demand prices equal to $SRMC$ in each case — an obviously nondiscriminatory outcome. The advantage of the discontinuous representation is in portraying the long-run solution, by permitting us simply to compare the vertical sum of the demand prices at any short-run capacity point like Q_A with the level of $LRMC$. In the continuous representation, we must still vertically sum the demand prices for the long-run solution, but since the effective short-run prices do not fall on a single vertical line, we cannot make use of the curve $D_1 + D_2$ for comparison with the level of $LRMC$. But the principle remains the same. In each case, in the short run the $SRMC$ for each class of demand is set equal to the demand price for the class; in the long run, the relevant $LRMC$ is to be compared with the sum of the effective demand prices (or, equivalently, with the sum of the $SRMC$'s).⁴

4. The following is an explicit discussion of the solution using the "traditional" continuous cost function of Figure II. In principle, the same two classes of cases occur as in the Figure I solution discussed earlier, but only the solution for the more interesting "shifting-peak" case, for which " $LRMC$ (joint)" is the relevant long-run marginal cost, will be given here. In the short run, suppose that the fixed investment is such as to dictate the short-run marginal cost curve

The danger which I believe Steiner fell into lay in letting the graphical convenience of the discontinuous representation lead him astray about the nondiscriminatory nature of the solution.

A word is perhaps desirable on the principle of charging according to short-run marginal cost, since the error of assuming that long-run marginal cost is the appropriate principle is all too common.⁵ In part this error seems to be due to an association of "short-run" with "short-sightedness" — failing to take account of the full consequences of one's actions. This is not the implication of short-run cost in economic theory; the concept takes into account *all* relevant costs, but under the assumption that one or more factors is fixed. A related error is to assume that since *SRMC* involves only variable costs, while *LRMC* involves both variable and capital costs, the latter must be greater.⁶ A glance at Figure II will show that *LRMC*

SRMC. The optimal quantities are Q^{AP} for the peak demand and Q^{AS} for the slack demand, the corresponding prices $Q^{APP'}$ and $Q^{AS'}$ being charged. The purchase of another unit of fixed plant can most simply be interpreted as involving a displacement of *SRMC* to the right by exactly the amount of "capacity" purchased (for a traditional cost function, the concept of "capacity" is not a technological datum but has to be given some arbitrary definition, which need not concern us). The *SRMC* curves have purposely been drawn with an initial horizontal phase at the level of c , and in the Figure II analysis c plays essentially the role which b played in the Figure I analysis. Since an increase in capacity shifts the entire *SRMC* curve to the right, the horizontal phase is stretched by the same amount—i.e., an increase in capacity permits a corresponding increase in the number of units which can be produced at a unit cost of c . Consequently, "*LRMC* (separable)" is at the level $c + \beta$, and "*LRMC* (joint)" at the level $2c + \beta$. In the present case, the equilibrium level of capacity generates the short-run curve *SRMC_B*, and the sum of the heights of the demand prices V' and T' equals $2c + \beta$. In the long-run solution therefore, the sum of the demand prices (and the sum of the short-run marginal costs along *SRMC_B*) equals the relevant long-run marginal cost — *LRMC* (joint). (The special assumption about the shape of the *SRMC* curves was only necessary to allow a simple graphical presentation.)

5. Both Davidson and Houthakker seem to commit this error. See Davidson, *op. cit.*, p. 72, and H. S. Houthakker, "Electricity Tariffs in Theory and Practice," *Economic Journal*, LXI (March 1951), 1-25. Westfield, in a review of Davidson's book, indicated that charging the short-run marginal cost was correct. See F. Westfield, *Current Economic Comment*, XVIII (Aug. 1956), 55-57. The question has also been discussed by Lewis, whose position is not entirely clear. He seems to indicate that prices equal to short-run marginal cost are efficient in the theoretical sense, but on the other hand he favors *LRMC* pricing — apparently on certain practical grounds. See W. A. Lewis, *Overhead Costs*, pp. 15-18.

6. Thus Davidson says that long-run marginal costs must be covered if firms are to stay in business, implying that covering short-run marginal costs would be insufficient (Davidson, *loc. cit.*). Similarly, Houthakker comments that existing electricity tariffs in Britain may not only be lower than *LRMC*, but may even be lower than *SRMC*, again implying that *SRMC* is necessarily lower (Houthakker, *op. cit.*, p. 10).

can perfectly well be less than $SRMC$ — the reason for expanding capacity is precisely the fact that new capital costs are incurred as a cheaper substitute for heavy variable costs which would have to be incurred if increased output were to be provided without an increase in the fixed factor ("capacity").

Now pricing according to $SRMC$ is appropriate for utilities because essentially all the contracts are of an extremely short-run nature. At any moment of time the utility has fixed capacity, and must dispose of its services to customers who can cancel out their demands instantly. Consequently, at any point the relevant costs are the short-run costs of meeting such demands. The utility, if it is wise, will increase capacity as long as $LRMC$ is below the relevant $SRMC$, so that in long-run equilibrium it will be the case that $SRMC = LRMC$ (in the peak load situation, the relevant $SRMC$ to equate with $LRMC$ is the sum of the $SRMC$'s incurred for those periods for which an expression of capacity will involve an increase of output). But at any moment of time, it can do no better (in a welfare sense) than to optimize in the short run by setting $SRMC$ equal to demand price.⁷ Thus, in Figure II with fixed investment such that $SRMC_A$ is appropriate, any attempt to charge the price OL equal to $LRMC$ will obviously be inefficient.

A word is perhaps appropriate on the question of recovery of total costs. In Steiner's analysis, since the levels of b and β are both constant, the long-run equilibrium prices just equal the total costs, and this total can be neatly segregated into variable and capacity costs. If we now retain all the other assumptions, but assume that β declines as a straight-line function of q , we see that the average capacity cost exceeds the marginal capacity cost. In this case, the equilibrium marginal conditions will fail to recover the total capacity cost. For other cases, the optimal "capacity" charges may exceed the capacity cost. Correspondingly, if we hold β constant but depart from the assumption that b is a horizontal straight line, the optimal "commodity" charges may exceed or fall short of the com-

7. Another reason given for basing charges on $LRMC$ is the fact that planning must be done well in advance (Davidson, *loc. cit.*). That $LRMC$ is relevant for planning is indubitable, but this is planning for capacity. The matter of pricing is equivalent to planning for current output, which is a short-run problem. Another practical consideration is that regulatory commissions are unlikely to permit the continuous price variation necessary to achieve continuous short-run optimization, but this does not affect the theoretical solution. Even as a practical matter, it is not completely impossible that practice of regulatory agencies (or of pricing administrators for government-owned utilities) might be affected by a demonstration of the relevance of $SRMC$ for efficient pricing.

modity costs. It is clear that any pricing principle based upon allocating or recovering total commodity or total capacity costs will not, except in special cases, be consistent with efficient prices in terms of the marginal conditions.

The whole controversy may be briefly summarized. Davidson, opposing discrimination and favoring the principle of pricing according to *long-run* marginal cost, ran aground on the peak load problem, where pricing on the *LRMC* principle leads to obviously wrong answers. Steiner, while providing the correct long-run solution, used a peculiar extreme idealization of the cost function which led him to infer that his solution involved discriminatory prices. I have shown, I believe, that using the usual cost function of economic theory, or with a deeper opportunity-cost interpretation of Steiner's cost function, we are led naturally to the short-run solution as well as to Steiner's long-run solution. The conclusion is that efficient utility pricing at any point in time should be based upon *short-run* marginal cost. A long-run welfare optimum involves, in addition, purchasing new capacity so long as the relevant *LRMC* is less than the sum of the *SRMC*'s for the "effective" periods in the peak load case, or is simply less than the *SRMC* if there were no peak load problem. (By "effective" periods we mean those for which the demand prices are sufficiently high so that an expansion of capacity will entail some increase in output.) In long-run equilibrium for the peak load case, then, the efficient prices remain equal to the respective *SRMC*'s, and their sum for the effective periods equals the relevant *LRMC*. Prices so based do not correspond with the fundamental idea underlying the concept of discrimination in economic theory, since the price differences all represent differences in the relevant social marginal costs of service — the only proviso being that marginal cost is to be interpreted in the opportunity-cost sense wherever a discontinuity in the cost function makes the usual cash-outlay marginal cost indeterminate.

FURTHER COMMENT

By H. S. HOUTHAKKER*

In a recent article Professor Peter O. Steiner¹ has criticized the solution for the peak load problem in electric power supply which I proposed in 1951.² He maintains that an optimal solution should involve different prices for equal outputs, whereas my proposal implies equal prices for different outputs. Thus in his example on page 594 my solution leads to equal prices for the actual peak period 1 and the potential peak period 2, a lower price being charged for the off-peak period 3, even though the use of capacity will then be less in period 2 than in period 3. Under his own solution prices would be such that the capacity used is the same in periods 1 and 2, which can in general be achieved only by having different prices in those periods, with a further different price in period 3.

It appears that Steiner has overlooked a constraint under which my solution was derived, namely that only two different prices are charged. Under that condition it is clear that the price must be the same for the actual and potential peak periods, for if period 2 were charged the same low price as period 3, demand in period 2 would exceed demand in period 1, and period 2 would become the actual peak period.

The reason why I introduced this constraint on the number of different prices was that it is costly to discriminate according to the time of consumption. Such discrimination requires special measuring devices, which are the more expensive the more different prices that can be charged. Accordingly my solution was intended mainly to determine the extent of the time interval to which each of the two prices applies, as well as the levels of those prices. I still think that, viewed in this light, my solution is correct and more realistic than Steiner's, which neglects the cost of discrimination.

The point is important for two reasons. In the first place the multiplicity of rate systems proposed by engineers during the last sixty years is not merely the result of confusion as regards the economics of the problem. It also reflects the fact that only a limited

* Stanford University

1. "Peak Loads and Efficient Pricing," this *Journal*, LXXI (Nov. 1957), 585-610.

2. "Electricity Tariffs in Theory and Practice," *Economic Journal*, LXI (1951), 1-25.

class of tariff types can be put into practice without prohibitive customer costs. Thus electric power companies have usually considered it impossible to base charges on the complete load curve of a consumer, since measurement of that curve would be too expensive. They have accordingly been content to divide customers into industrial, residential, commercial, etc., under the assumption that this classification will to some extent characterize their demand over time. One purpose of my 1951 paper was to argue that a more accurate solution, involving discrimination according to time of consumption, was feasible despite the cost of special meters.

The second aspect is of more general importance. Much of the sterility of modern welfare economics can be attributed to the neglect of institutional and technological constraints which in reality make the range of possible solutions much narrower than those theories would have us believe. If almost everything is possible, it is not surprising that there can be no consensus as to which arrangements are optimal. A large part of contemporary welfare economics consists of ponderous proofs of the impossibility of such a consensus. When it is recognized that there are additional constraints the whole problem becomes much more determinate and the scope for individual idiosyncrasies is reduced. If welfare economics is ever to be of assistance to economic policy, it will have to pay more attention to such constraints. It is a step backwards if realistic solutions are condemned because they do not agree with the unduly general criteria of abstract theory.

REPLY

By PETER O. STEINER*

Professor Houthakker's 1951 paper made a major contribution to our understanding of the peak load problem in electricity supply by showing that a two-rate time-of-day tariff produced a better use of resources than a tariff based upon a single daily rate, albeit at some increase in customer costs. Whether introduction of the time of day tariff is justified depends upon whether the gains exceed the added costs. He decided that they probably would.

Whether three rates would be better than two involves precisely the same sort of comparison. I attempted to show that, under certain circumstances, the optimal number of rates might be greater than two. Houthakker does not consider anything beyond a two-rate system because (1) "The introduction of further rates yields rapidly diminishing returns while consumer costs increase accordingly," and (2) for the separation of energy and capacity costs "two rates appear sufficient."¹

The first reason is an assertion about what is clearly an empirical matter. Rapidly diminishing benefits are a priori neither necessary nor obvious, either in electricity supply or in a number of other fields with peak load problems. On the cost side, with respect to electricity, the major equipment cost appears to be in identifying a number of periods in the day. Indeed one of the devices Houthakker suggests, the clock-meter, would appear to have diminishing marginal costs per period identified after the second. In any case it is not apparent that the four period, two-rate scheme suggested by Houthakker would have much cost advantage over (say) a three-rate scheme for the same four periods. But I do not propose to prejudge the empirical question in electricity or elsewhere. To identify the nature and number of rates required for optimal resource use seems worthwhile in order that the marginal benefits and costs may be compared.

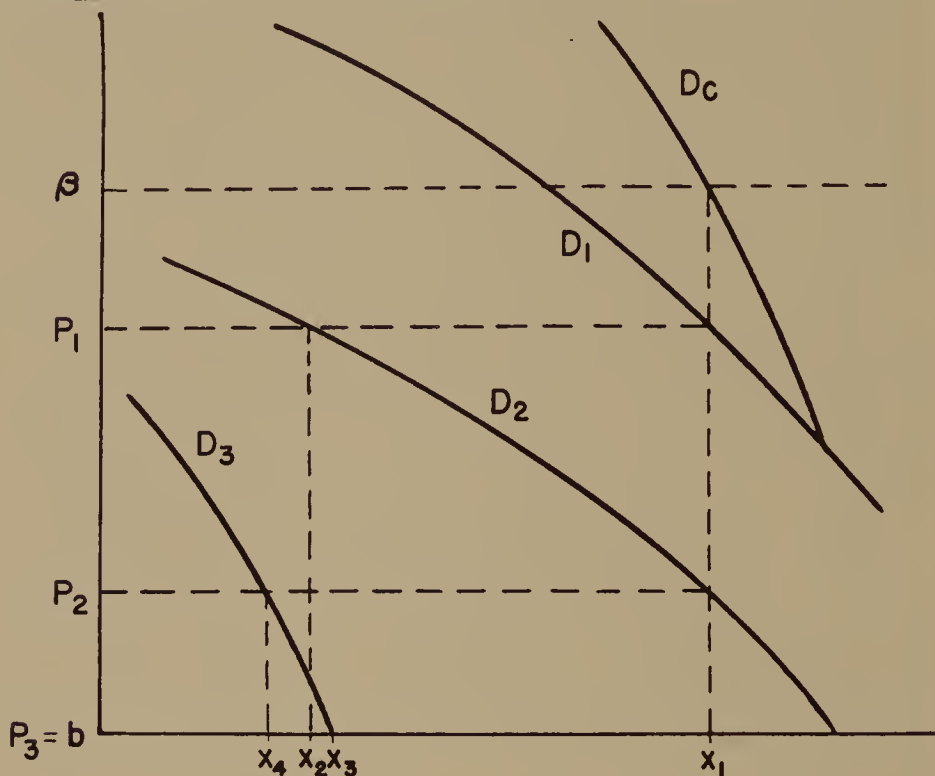
The second reason given is, in general, either trivial or wrong. Houthakker's two rates are not in general sufficient to secure an optimal amount of capacity or optimal use of that capacity, as I have shown. That two rates are both necessary and sufficient to secure *some* distinction between energy and capacity costs is hardly surprising. It is not even clear that his suggested rates are the two best rates.

Consider the accompanying figure (similar to my earlier Figure II), which is constructed so that $p_1x_2 = p_2x_1$ and thus Houthakker's

*University of Wisconsin.

1. Houthakker, *op. cit.*, 17.

two rates lead to the optimal total capacity. He would charge P_1 in periods 1 and 2 and P_3 in period 3, with quantities x_1 , x_2 , and x_3 respectively. The nonoptimality of this scheme lies in its underutilization of the existing capacity in period 2. In what sense is



this scheme superior to the following *two* rate scheme: Charge P_1 in period 1 and P_2 in periods 2 and 3? The latter, as drawn, generates *more* revenue, causes heavier use of the *same* capacity (since $x_1 - x_2 > x_3 - x_4$), and incidentally is accompanied by more consumers' surplus. It does to be sure provide revenues in excess of cost by soaking the period 3 users, but we are constrained by the two price limit not to lower the cost to these users to P_3 . If the excess revenue is regarded as undesirable we can, of course, lower slightly both P_1 and P_2 and substitute socially 'unjustified' capacity for the excess revenue. But is either of these worse than the underutilization of capacity of Houthakker's scheme?²

2. In other words we pay a price for the limitation to two prices. The Houthakker prices are best only if we assume some additional constraints, e.g.: (1) the *lower* rate be equal to operating costs, and (2) given constant costs, we precisely recover total costs. I know of no technological or institutional restraints which make these particular assumptions compelling. As Hirshleifer shows above total cost recovery is due to the assumed constancy of costs and is not an inherent feature of any of the pricing schemes under discussion.

Finally let me note that while I did criticize Houthakker's solution I did not condemn it. It represented both theoretically and practically an important step forward. But to be blinded by particular assumptions as to what is or is not technologically or institutionally feasible seems unfortunate not only because it prejudges factual matters³ but also because both technology and institutions do vary and change, and it is perhaps worth knowing the opportunities foregone because of a particular (perhaps incorrect, perhaps remediable) constraint.

Professor Hirshleifer's paper represents a notable addition to the discussion. I agree with most of it and appreciate his explicit discussion of the short-run optimal solution and his discussion of the limiting assumptions required in order that "efficient" pricing lead to recovery of total costs.

We differ, first, as to whether discontinuous marginal costs are an idealized situation adopted for graphic convenience (his view) or whether they are, for the planning problem I addressed a fundamental (and by the way mathematically inconvenient) consequence of the appropriate total cost function. My view was expressed briefly in the original paper (footnote 9, p. 590) and I shall not pursue it further.

The second difference concerns whether the optimal solution (about which we agree) is discriminatory. Since this issue does not affect the solution it would be both painless and gracious to accept Hirshleifer's definition of discrimination and to concede that the optimal solution is nondiscriminatory. His is a sophisticated definition and an intriguing one. It may even be right.

I confess to some reservations. Marginal cost, Hirshleifer argues, should be interpreted as opportunity cost, the most valuable foregone alternative, including opportunities foregone in the same market. Under this interpretation it seems to me that (1) it is generally impossible to define marginal cost for a product independent of the demand function for that product, and (2) profit-maximizing marginal cost is in general always equal to (less only an epsilon) or *above* price.⁴ Has this definition saved the concept of price discrimination by annihilating the concept of marginal cost?⁵

3. Indeed it seems to me that it is the multiplicity of possible judgments as to what is feasible rather than the absence of such judgments that most often produces a multiplicity of solutions and a lack of consensus.

4. It is above price, for example, in the lower priced market if we have the "textbook" type of price discrimination that Hirshleifer cites.

5. In what I suppose to be the usual view, marginal cost is the derivative of a total cost function where costs reflect not cash outlays but the opportunity costs of the factors used. Often, if not always these opportunity costs are reflected

Both issues between us, then, rest upon the correct interpretation of marginal cost, and thus ultimately upon the nature of the total cost function. My long-run cost function,

$$\text{Total cost} = b \sum_i x_i + \beta \max_i (x_i)$$

is perhaps too simple in many ways (e.g., in assuming b and β to be constants rather than functions of output) but at least it is explicit. We might readily settle our differences if Hirshleifer would make his total cost function explicit.

in factor prices, actual or imputed. By excluding from opportunity cost the unsatisfied demand for the product it is possible to define the cost function independently of the demand function and I for one have found it convenient not only to describe maximizing behavior in terms of marginal conditions (e.g., marginal revenue = marginal cost) but also to make something of such things as the divergence of price and marginal cost. If we adopt the notion that marginal cost is never less (for a continuous demand curve) than price (minus epsilon) much is lost.

But even if we follow Hirshleifer and let the demand of the marginally unsatisfied customer be included in the opportunity cost, I am uncertain why the appropriate marginal opportunity cost is not the change in total revenue of a shift to the next most valuable customer. This, of course, would correspond to *marginal* revenue, rather than price, and would leave the issue of the presence or absence of discrimination dependent upon the elasticities of demand in the two periods

INVESTMENT CRITERIA AND CAPITAL INTENSITY ONCE AGAIN

By ALBERT O. HIRSCHMAN AND GERALD SIRKIN*

We hesitate to add more pages to the discussion of the article on investment criteria by Galenson and Leibenstein.¹ But we feel that the basic point which in the original article was either in error or obscure has not been clarified and that, as a consequence, the real contribution which it contained has not yet been brought out.

1. One method of allocating investment is to choose the investment pattern that would be selected by the market mechanism. Alternative investment criteria would presumably produce an allocation different from that chosen by the market. Galenson and Leibenstein have proposed a "reinvestment" criterion: investment should be allocated so as to maximize future levels of saving and, thereby, future levels of investment.

2. In the Galenson-Leibenstein model, output is divided between wages and profits. Wages are entirely consumed and profits are all saved.² In such circumstances, saving would obviously be maximized by maximizing profits. But, as we all know, the market itself maximizes profits. Therefore, no case is made for interfering with the market mechanism and the reinvestment criterion is superfluous *as long as we assume that profits are entirely reinvested*. On this assumption, a planning agency would reduce profits and, hence, reinvestment if it were to favor projects and techniques that are more capital-intensive than those that would be selected by the market.

3. To establish reinvestment as an independent criterion, we must discard the assumption that all profits are reinvested. It is a

* Yale University

1. Cf. W. Galenson and H. Leibenstein, "Investment Criteria, Productivity and Economic Development," this *Journal*, LXIX (Aug. 1955), 343-70 and discussion by H. Neisser, J. Moes, H. H. Villard, and the original authors in the issues of November 1956, February 1957, and August 1957. The articles by O. Eckstein and A. K. Sen in the February and November 1957 issues of this *Journal* also deal with related problems.

2. The assumption that all profits are reinvested appears in the "model" of the original article (pp. 357 and 359). In private correspondence, Leibenstein has commented to us that there was no intention to establish a firm connection between this assumption and the reinvestment criterion. Indeed, in their reply to Moes (p. 473), Galenson and Leibenstein give an example in which the assumption is discarded. It is the purpose of our note to show that (1) the assumption *must* be given up if the reinvestment criterion is to have an independent existence; and (2) precisely by giving it up one can rebuild the argument in favor of capital intensity in such a manner that it gains in solidity what it loses in sweep.

particularly unrealistic one for underdeveloped countries where "milk-ing" of enterprises is frequently the rule and "ploughing back" the exception. If entrepreneurs consume and if some types of entrepreneurs have higher propensities to consume than others, then reinvestment may become a genuine criterion, leading to choices different from those resulting from the market mechanism. For then a project *A*, while less profitable than a project *B*, may yield more saving, provided only the owners of *A* save more out of their (smaller) profits than the more affluent, but also more profligate, owners of *B*. In this case, the reinvestment criterion would choose *A* while the market would choose investment in *B*.

4. Whether reliance on the reinvestment criterion rather than on the market would lead to more capital-intensive investment is now seen to depend on whether capital-intensive capitalists as a group have a higher propensity to save than other types of property owners. In general, a plausible case can be made for the presumption that they do. The threat of obsolescence and the attractiveness of new and better machines make the capitalists with expensive machinery more accumulation-minded than entrepreneurs with relatively little capital input. Much has been written about the difference in this respect between landlords and industrialists. But it is perhaps more justified to distinguish between all capital-intensive property owners on the one hand and all labor- or land-intensive property owners on the other. There is, for instance, evidence that the land-intensive hacienda owner is far less of a saver than the modern capital-intensive plantation owner.³ The way in which capital intensity induces the ploughing back of profits has much in common with the manner in which certain types of modern technology make for efficient management, better co-ordination of tasks, and higher labor productivity.⁴

5. The emphasis on reinvestment can, therefore, yield an argument for investment in more capital-intensive industries than would be indicated by a pure market calculation. The weight that should be given to the argument is a matter for empirical investigation in each situation and country. If the propensity to save out of profits is at all a function of the degree of capital intensity, it is likely to be a step function rather than a continuously increasing one. Therefore, the influence of the reinvestment consideration will probably

3. E. R. Wolf and S. W. Mintz, "Haciendas and Plantations in Middle America and the Antilles," *Social and Economic Studies*, VI (Sept. 1957), 380-412.

4. A. O. Hirschman, "Investment Policies and 'Dualism' in Underdeveloped Countries," *American Economic Review*, XLVII (Sept. 1957), 563 ff.

be more pertinent to the choice among industries with widely differing capital intensities than to the choice between more or less capital-intensive techniques within an industry.

6. The implications of the Galenson-Leibenstein thesis for another investment criterion, social marginal productivity, are also worth noting. In many underdeveloped countries, the actual wage rate paid in industry exceeds the marginal productivity of labor in agriculture or services. In such cases the SMP criterion would involve assigning to labor a "shadow" or "accounting" wage rate lower than the prevailing rate. The effect of using such accounting wage rates as guides to allocation would be to make investments more labor-intensive and total output greater than they would be if market wage rates were used; but total profits would be smaller since hiring decisions would have been made on the basis of wage rates different from those which firms must actually pay. It follows that, if reinvestment is maximized by maximizing profits, the reinvestment criterion in effect re-enthrones market prices as the best guides for investment decisions, even in cases where these prices do not correctly reflect social costs.

THE ROCKEFELLER REPORT: IMPLICATIONS FOR FISCAL AND MONETARY POLICY

By ARTHUR SMITHIES*

The "*Rockefeller Report*"¹ on the United States economy is an important document from the point of view of both content and method. I shall not dwell on the fact that a group sponsored by the Rockefeller brothers and consisting of members who are not inclined to exaggerate the role of government, have produced forthright estimates of rapidly expanding government requirements for defense and nondefense purposes over the next decade. My concern here is with the implications of the Report for the methods to be used in determining fiscal and monetary policy.

The feature of the Report that deserves special attention is the use of a ten year projection of the expenditure requirements of all levels of government as the starting point for the determination of current policy.² This is in sharp contrast to traditional practice. The federal budget has rarely been projected for more than one year in advance, and the Employment Act of 1946 is more concerned with employment and purchasing power in the immediate future than with the longer run.

The Report estimates, on the basis of studies by other Rockefeller panels, that over the decade total government expenditures should rise from a level of \$114 billion in 1957 to a "low" figure of \$171 billion in 1967, or to a "high" figure of \$203 billion. For the purpose of illustrating the method, I shall use the mean figure of \$187 billion. The corresponding figures for government expenditures on goods and services are \$86 billion in 1957 and \$140 billion in 1967.

The Report next considers private consumption. It assumes, in effect, that a viable policy requires that consumption provision must be made not only for a rising population, but for expectations of higher living standards. In recent years per capita consumption has grown at 2 per cent a year. If that trend is to be continued for

* Harvard University.

1. *The Challenge to America: Its Economic and Social Aspects*, (New York, 1958).

2. The Report owes much to the pioneer work of Gerhard Colm, and particularly to his *The American Economy in 1960* (written in 1952). But the emphasis is different. Whereas Colm's work is concerned principally with alternative methods of maintaining full employment, the Rockefeller Report addresses itself to the attainment of specific goals with respect to government expenditure programs and consumption under conditions of full use of productive capacity.

the next decade, consumption should increase from \$281 billion in 1957 to about \$400 billion in 1967.

Thus government goods and services and private consumption together should increase from \$367 billion to \$540 billion. For the sake of simplicity, let us assume that these figures are points on an exponential growth curve. In that event the two items together must grow at an annual rate of about 4 per cent a year.

Assuming that the economy does not run into labor shortages (or chronic unemployment), the problem then is to ensure that capital formation will be sufficient to attain the 4 per cent rate. Policy must be designed to achieve both the required gross saving rate and the required inducement to invest.

First, with respect to the saving ratio: the record of the past indicates that a 3 per cent growth rate of GNP has been associated with a 15 per cent gross saving ratio. This implies a marginal gross capital-output ratio of 5.³ To achieve a 4 per cent growth rate, a 20 per cent gross saving ratio is therefore needed. Long-run policy must therefore increase the saving ratio from the present 15 per cent to 20 per cent. This could be done by additional general taxation.

But, at the same time, the inducement to invest must be strengthened, so that businesses have an incentive to invest 20 per cent of the GNP. Hence the general tax increases designed to increase saving (private *plus* government) must be tempered with special investment incentives such as relief from profits taxation and allowance of accelerated amortization, or by an easier credit policy. This is the course of action the Report recommends.

But other alternatives may be available. First, a basis now exists for reviewing the government estimates in the light of their economic implications. The government, rightly or wrongly, may decide that some of the proposed programs are not worth the effort involved in accelerating the rate of growth.

Second, the consumption target should be reconsidered. A 1 per cent growth rate may be sufficient to satisfy rising expectations; and in that case the economy could continue on its present course. Or improvement in living standards may be consistent with more economical use of durables. For instance, special excise taxes or consumer credit controls could be used to induce the public to use their durables for a longer part of their useful life.

3. This gross concept must not be confused with the net ratio which is, of course, much lower. The use of the gross ratio is obviously a crude first approximation since it assumes implicitly that depreciation and obsolescence of existing equipment is uniquely related to the desired increase in capacity.

With respect to short-run policy, the Report proposes that taxes be reduced to relieve the present recession and be subsequently increased along the lines required by long term needs. Here I disagree. A general reduction, unless explicitly of a temporary character, is very likely to become permanent. The long-run picture suggests that we should not run the danger of reducing the tax base. Relief from the recession can be obtained by rapid introduction of some of the new expenditure programs the Report considers necessary.

Objections to this entire approach may be raised that our knowledge of both government requirements and the structure of the economy is so uncertain that a ten year projection cannot feasibly be used as a basis for current policy. Such arguments, however, should lead not to concentration on the short run but to recognition that the long run is uncertain and to the making of appropriate strategic assumptions. If, for instance, we err on the high side in estimating defense requirements, subsequent reduction is relatively easy. The problem is of a different order of difficulty when estimates are too low and the economy must be made to accommodate increases in the defense budget.

The specific figures in the Report clearly require much further discussion; and the economic analysis needs to be defined. But even if major revisions turn out to be needed, the Report has done invaluable service in providing a framework for the consideration of policy and an agenda for research.

RECENT PUBLICATIONS

- AMERICAN TARIFF LEAGUE, INC. *The United States in World Trade.* New York. 1958. pp. 113.
- ARROW, KENNETH J.; KARLIN, SAMUEL; AND SCARF, HERBERT. *Studies in the Mathematical Theory of Inventory and Production.* Stanford, California: Stanford University Press. 1958. pp. 340. \$8.75.
- BOLLES, BLAIR. *The Big Change in Europe.* New York: W. W. Norton and Company. 1958. pp. 527. \$5.95.
- BUCHANAN, JAMES M. *Public Principles of Public Debt.* Homewood, Illinois: Richard D. Irwin, Inc. 1958. pp. 223. \$5.00.
- CHAMBERLIN, EDWARD H.; BRADLEY, PHILIP D.; REILLY, GERARD D.; POUND, ROSCOE. *Labor Unions and Public Policy.* Washington, D. C.: American Enterprise Association. 1958. pp. 177. \$4.50.
- DE LA VEGA, JOSEPH. *Confusion de Confusiones.* Boston, Massachusetts: Harvard Graduate School of Business Administration. 1957. pp. 41.
- DUESENBERY, JAMES S. *Business Cycles and Economic Growth.* New York: McGraw-Hill Book Company, Inc. 1958. pp. 341. \$6.50.
- FFORDE, J. S. *An International Trade in Managerial Skills.* Oxford, England: Basil Blackwell. 1957. pp. 153.
- GALBRAITH, JOHN KENNETH. *The Affluent Society.* Boston, Massachusetts: Houghton Mifflin Company. 1958. pp. 368. \$4.75.
- GHOSH, O. K. *Problems of Economic Planning in India.* Allahabad, India: Kitabistan Publishers. 1957. pp. 159.
- GORDON, WENDELL C. *International Trade: Goods, People, and Ideas.* New York City: Alfred A. Knopf, Inc. 1958. pp. 647. \$6.75.
- HAMILTON, DANIEL C. *Competition in Oil.* Cambridge, Massachusetts: Harvard University Press. 1958. pp. 233. \$6.75.
- HICKMAN, W. BRADDOCK. *Corporate Bond Quality and Investor Experience.* Princeton, New Jersey: Princeton University Press. 1958. pp. 536. \$10.00.
- LERNER, MAX. *America as a Civilization.* New York: Simon and Schuster. 1957. pp. 1035. \$10.00.
- McFARLAND, DALTON E. *Management Principles and Practices.* New York: The Macmillan Company. 1958. pp. 612. \$6.95.
- NEF, JOHN U. *Cultural Foundations of Industrial Civilization.* New York: Cambridge University Press. 1958. pp. 164. \$4.00.
- PEACOCK, ALAN T.; TURVEY, RALPH; STOLPER, WOLFGANG F.; HENDERSON, ELIZABETH (EDITORS). *International Economic Papers No. 7.* New York: The Macmillan Company. 1958. pp. 184. \$3.50.
- ROYAL COMMISSION ON CANADA'S ECONOMIC PROSPECTS. *Final Report.* Ottawa, Canada. 1957. pp. 509.
- SCHLOSS, H. H. *The Bank For International Settlements.* Amsterdam: North-Holland Publishing Company. 1958. pp. 184. \$4.00.
- TURVEY, RALPH AND WISEMAN, JACK (EDITORS). *The Economics of Fisheries.* Rome: FAO of United Nations. 1957. pp. 234.
- UNITED NATIONS. *1957 Economic Survey of Europe.* New York: Columbia University Press. 1958. \$4.00.
- . *Industrialization and Productivity.* New York: Department of Public Information, United Nations. 1958. pp. 78. 70 cents.
- . *Integrated River Basin Development.* New York. 1958. pp. 60. 70 cents.
- . *Statistical Yearbook 1957.* New York: Columbia University Press. 1957. pp. 674. \$6.50.
- U. S. STEEL CORPORATION. *Steel and Inflation — Fact vs. Fiction.* New York: U. S. Steel Corporation Public Relations Department. 1957. pp. 292.

ÉCONOMIE APPLIQUÉE

Tome XI

JUILLET-SEPTEMBRE, 1958

No. 3

THEORIE MONETAIRE

F. J. DE JONG
A. GAMBINO
D. HOROWITZ
M. COHEN
L. DAVIN
M. BYE
R. BERTRAND

La vitesse de circulation de la monnaie et l'équilibre monétaire.
L'offre de crédits.
Le marché des capitaux et la politique monétaire.
Dépenses de consommation, épargne et avoirs liquides.
Note sur la dynamique du déséquilibre des changes.
Le rôle du capital dans le développement économique.
Comparaison des tarifs douaniers des Pays du Marché Commun (Cahier 64).
Enquête sur les zones moins développées de l'Alsace (Cahier 65).
Le développement économique, analyse et politique (Cahier 66).
Planification indicative et développement économique (Cahier 67).
Sur la croissance économique des pays sous-développés (Cahier 68).
L'économie régionale, espace opérationnel (Cahier 69).

J. L. REYNAUD
R. BARRE
J. BENARD et J. W. HACKETT
G. DESTANNE de BERNIS
J. R. BOUDEVILLE

Economie Appliquée et Cahiers de l'Institut de Science Economique Appliquée
Directeur: François Perroux Secrétaire Général: H. Bonjour
(I.S.E.A., 35 Boulevard des Capucines, Paris — 2ème)

REVUE DE SCIENCE FINANCIÈRE

THE FRENCH REVIEW OF PUBLIC FINANCE

Number III. July-September 1958

Budget Estimating in the International Organizations
Barriers to the Growth and Organization of Markets
in Under-developed Countries
The Corporation Tax as an Instrument of Economic Policy
Handicrafts and Taxation
Financial and Economic Aspects of the Concept
of "Impasse" (Budget gap)
Survey of French Finance
Survey of Tax French Legislation
Book Reviews

C-A. Colliard (Grenoble)
L. de Carbon (Nancy)
L. Dini (Florence)
M. Penouil (Bordeaux)
J-L. Vauzanges (Paris)

Bibliography
French and International Financial Documents

The Review is published quarterly. Yearly subscription abroad 4,000 French francs or 9.50 U.S.A. dollars. Orders can be placed to Pichon and Durand-Auzias, 20, rue Soufflot, Paris (V°) and through Stechert-Hafner, 31 East 10th Street, New York, N. Y.

METROECONOMICA

INTERNATIONAL REVIEW OF ECONOMICS

Number 2

Volume X

A letter from Vilfredo Pareto to Irving Fisher.
Business investment and the marginal cost.
Mr. Armstrong and the theory of demand.
A restatement of the conditions for identifiability
in complete systems of linear difference equations.
The problem of intrinsic complementarity and separability of goods.
The valuation of random income streams.
A macro-economic model of growth and fluctuation.
On parameter estimates provided by various methods.

N. J. Cunningham
H. A. J. Green
H. S. Konijn
M. Morishima
A. Roy
A. Ll. Wright
A. Zellner

Board of Editors: C. Bresciani Turrone, L. Dupriez, E. Fossati (*Managing Editor*), R. Frisch, W. Leontief, G. Lufalla, A. Mahr, J. Marschak, G. U. Papi, R. Roy, E. Schneider, G. L. S. Shackle, R. Stone, J. Tinbergen, G. Tintner, F. Zeuthen.
Published every four months. Subscription rate \$7.00 per year (two years \$12.00). Communications to the editor to be addressed to *Professor Eraldo Fossati*, director of the Laboratorio di economia Vilfredo Pareto, *Università, Genova*, (Italy). Communications regarding subscriptions should be sent to *Licinio Cappelli*, Corso 12, Trieste or *G. E. Stecher-Hafner Inc.*, 31 East Tenth Street, New York 3, N. Y.

THE JOURNAL OF POLITICAL ECONOMY

Volume LXVI

June 1958

Number 3

How Income Ought to be Distributed: A Paradox in Distributive Ethics
Liquidity and Farm Investment Outlays
New Developments on the Oligopoly Front
Futures Trading and the Storage of Cotton and Wheat
Underemployment in Backward Economies — Some Additional Notes
Underemployment in Backward Economies: An Empirical Comment
Reply
Book Reviews and Books Received

Robert H. Strotz
Norman J. Breckner
Franco Modigliani
Lester G. Telser
Harvey Leibenstein
Harry T. Oshima
Harvey Leibenstein

Edited by Albert Rees in Cooperation with the Other Members of the
Department of Economics of the University of Chicago

FOUNDED IN 1892. Published bimonthly: February, April, June, August, October, December. Subscription rate \$6.00 per volume. Single copy \$1.50. Special rates: \$5.00 per volume for faculty members teaching business or economics, \$3.00 per volume for graduate students in business or economics. Canadian postage 40 cents. Foreign postage 75 cents.

The University of Chicago Press, 5750 Ellis Avenue, Chicago 37, Ill.

American Economic Review

CONTENTS

VOLUME XLVIII

JUNE 1958

NUMBER 3

ARTICLES

Cost of Capital and the Theory of Investment

Franco Modigliani and M. H. Miller

Policy Analysis

C. E. Lindblom

Capital Longevity and Economic Development; Mathematical Note on

Optimum Longevity

R. C. Blitz and F. M. Westfield

"People's Capitalism" and Stock-Ownership

Victor Perlo

The Production Ceiling and the Turning-Point

K. D. Roose

Ricardo and the 93 % Labor Theory of Value

G. J. Stigler

REVIEW ARTICLES

Advances in Game Theory

H. M. Wagner

The Dollar Shortage Re-Visited

C. P. Kindleberger

The Permanent Income Hypothesis

H. S. Houthakker

COMMUNICATIONS

The Mutual Influence of Mitchell and Commons

Joseph Dorfman

The Union and Wages in Basic Steel: A Comment

Lloyd Ulman

Reply

Albert Rees

The AMERICAN ECONOMIC REVIEW, a quarterly, is the official publication of the American Economic Association and is sent to all members. The annual dues are \$6.00. Address editorial communications to Dr. Bernard F. Haley, Editor, American Economic Review, Stanford University, Room 220, Stanford, California; for information concerning other publications and activities of the Association, communicate with the Secretary-Treasurer, Dr. James Washington Bell, American Economic Association, Northwestern University, Evanston, Illinois. Send for information booklet.

For the First Time in English

A complete translation of the leading monthly Soviet Journal

ВОПРОСЫ ЭКОНОМИКИ PROBLEMS of ECONOMICS

Voprosy Ekonomiki is the journal of the U.S.S.R. Academy of Sciences, Institute of Economics. Theoretical and descriptive articles on all phases of capitalist and socialist economic life. An indispensable source for all thoughtful students of contemporary affairs.

AMONG THE CURRENT ARTICLES:

- The Problem of Catching up with the U.S. in Per Capita Production of Meat, Milk and Butter
- Result of Reforming the System of Economic Management
- The Law of Value and Problems of Price Formation in the U.S.S.R.

\$50 per year (12 issues)

\$4.50 per single copy

\$20 to individuals associated with institutions already subscribing

INTERNATIONAL ARTS AND SCIENCES PRESS

Authorized Translator and Publisher

33 West 42nd Street

New York 36, N.Y.

SELECTED PAPERS IN ECONOMIC THEORY

By **KNUT WICKSELL**. *Edited with an Introduction by* **ERIK LINDAHL**. This book makes available to the English reader some of the most important and interesting papers by Knut Wicksell. His main theories — on the value of money, marginal productivity, the nature of capital, etc. — are presented in a very readable form, while an introductory essay by Erik Lindahl tells the story of his life and gives some insight into his forceful personality. \$6.50

FRENCH BANKING STRUCTURE AND CREDIT POLICY

By **J. S. G. WILSON**. The first complete study of this subject in English, **FRENCH BANKING STRUCTURE AND CREDIT POLICY** gives detailed accounts of the French banking system, the various institutions that form the money market, the evolution of the Bank of France as the central bank, and the structure of the present hierarchy of authority. Mr. Wilson shows how the system of credit controls developed since 1945 has strengthened the Bank of France and emphasizes some additional steps that need to be taken to maintain economic stability. \$8.50

THE NIGER AGRICULTURAL PROJECT An Experiment in African Development

By **K. D. S. BALDWIN**. In recent years many heavily capitalized schemes have been started to assist under-developed areas. Although wide experience has been utilized many of them have failed to achieve their objects.

Mr. Baldwin has made a critical socio-economic analysis of the history of a particularly well-documented agricultural project in Nigeria. Here, he discusses the handling of the problems arising in the initial planning stages of the project and the causes of its failure. 4 maps. 8 pages of illustrations. Tables. \$5.00



Through your bookseller, or from

HARVARD UNIVERSITY PRESS

79 Garden Street, Cambridge 38, Massachusetts

**The best service
at the lowest cost
in the long run
depends on good earnings**



Low telephone earnings do not mean low rates
Good telephone earnings do not mean high rates

Many years ago the Bell System
pledged itself to provide the best
possible service at the lowest possi-
ble price.

We meant it then and we mean
it now.

Today, more than ever, it is evi-
dent that the best service at the low-
est cost in the long run depends on
good earnings.

To a considerable extent the pub-
lic, and we are afraid many who
should know better, have come to
think that low earnings mean low
rates and good earnings mean high
rates.

Yet few people have the idea that
the lowest earning soap company
makes the best and cheapest soap.
Or the lowest earning meat packer
makes the best and cheapest hams.
Or that the lowest earning company
in any line makes the best and
cheapest products and renders the
best service.

It doesn't apply to the telephone
business either.

There are many ways in which
telephone users benefit in both the
cost and quality of service through
good earnings for the telephone
company.

BELL TELEPHONE SYSTEM





The McGraw-Hill
Economics Handbook Series

Burns — SOCIAL SECURITY AND PUBLIC POLICY

300 pages, \$5.90

**Duesenberry — BUSINESS CYCLES AND ECONOMIC
GROWTH**

335 pages, \$6.50

Hansen — A GUIDE TO KEYNES

237 pages, \$4.90

Hansen — MONETARY THEORY AND FISCAL POLICY

236 pages, \$4.90

Hansen — THE AMERICAN ECONOMY

199 pages, \$5.00

**Harris — INTERNATIONAL AND INTERREGIONAL
ECONOMICS**

564 pages, \$7.00

Henderson & Quandt — MICROECONOMIC THEORY

Ready in September

Hoover — THE LOCATION OF ECONOMIC ACTIVITIES

310 pages, \$5.50

Kindleberger — ECONOMIC DEVELOPMENT

325 pages, \$6.50

Lerner — ECONOMICS OF EMPLOYMENT

397 pages, \$6.00

**Mikesell — UNITED STATES ECONOMIC POLICY AND
INTERNATIONAL RELATIONS**

341 pages, \$5.50

Schelling — NATIONAL INCOME BEHAVIOR

291 pages, \$5.50

McGRAW-HILL BOOK COMPANY, INC.

330 West 42nd Street

New York 36, N. Y.

Send for
copies on
approval

PRINCETON UNIVERSITY PRESS

As Unions Mature

An Analysis of the Evolution of American Unionism

By Richard A. Lester

As Unions Mature is a forcefully-written book about how American unions got to be what they are, where they stand now, and where they go from here. Richard A. Lester, Princeton University economist and expert on industrial relations, suggests that the missionary zeal and anti-boss militancy have gone out of unions as they have grown older, larger, more powerful and more settled. He tells how union ideals tend to break down, how democracy in unions tends to collapse, and how corruption can creep in. The author's objective appraisal gives a foundation for understanding mature unions of today and for intelligent judgment of current proposals for union reform.

184 PAGES. \$3.75

The French Economy and the State

By Warren C. Baum

This book provides a full view of the French economy and how it has affected France as a state. The author deals with government efforts to reconstruct the economic system and control inflation and recession cycles, and shows how the state acts as regulator of private enterprise in industry and agriculture. He also discusses the state's traditional role of public financier and its less orthodox role as entrepreneur in a large part of the economy now under public ownership. Mr. Baum's extensive use of original French source material and discussions with French officials in business and government make this book a unique contribution to understanding the economic and political problems of modern France. *A RAND Corporation Research Study.*

404 PAGES. TABLES. \$7.50

A Critique of the United States Income and Product Accounts

Studies in Income & Wealth, Volume 22

By the Conference on Research in Income & Wealth

The first systematic and critical analysis of the Office of Business Economics Estimates, including an examination of many of the theoretical issues in national income accounting and a discussion of the income and expenditure sides of the accounts. Papers appraise the suitability of the data for short-term analysis and the problem of obtaining *real* natural product in natural income measurement. *Published for the National Bureau of Economic Research.*

600 PAGES.
TABLES & CHARTS. \$11.50

Order from Your Bookstore, or

Princeton University Press, Princeton, N. J.

THE INTERNATIONAL ECONOMY

Revised Edition by P. T. ELLSWORTH, University of Wisconsin

Combining a presentation of basic economic principles applicable to international economic problems with an historical survey of the changing structure of international relations, this revised edition has been thoroughly rewritten to incorporate recent developments. Such current problems as European disequilibrium, the development of backward areas, and the reconstruction of world trade are thoughtfully discussed.

1958, 513 pages, \$6.75

MANAGEMENT PRINCIPLES AND PRACTICES

by DALTON E. McFARLAND, Michigan State University

This up-to-date work on general managerial functions integrates practice with theory and management functions with human relations and social science research. Study questions, case materials, and reference guides are included.

1958, 612 pages, \$6.95

HOUSING: A Factual Analysis

by GLENN H. BEYER, Cornell University

This comprehensive treatment of housing covers the demand and supply factors of the market; production, financing and tenure; design criteria; and environmental factors. Charts, sketches and maps supplement the text.

1958

355 pages

\$6.75

ESSENTIALS OF ACCOUNTING

by WILLIAM A. PATON and ROBERT L. DIXON, University of Michigan

1958, 800 pages, \$7.50

PROBLEMS AND PRACTICE SETS

by WILLIAM A. PATON, JR., ROBERT L. DIXON, and ROBERT H. COJEEN

Summer 1958

RULED FORMS

1958, paper, \$4.50

TEACHERS GUIDE available gratis

INTERNATIONAL ECONOMIC PAPERS, No. 8

The contributors: Ragnar Bentzel, Maurice Byé, Francesco Ferrara, Rumen Janakieff, Leif Johansen, Erik Lundberg, Jurg Niehans, Erich Schneider, Heinrich von Stackelberg, Stefan Varga, Alfred Weber.

Fall 1958

The Macmillan Company

60 FIFTH AVENUE, NEW YORK 11

THE
QUARTERLY
JOURNAL OF
ECONOMICS

FOUNDED 1886

Vol. LXXII November, 1958 No. 4

TRAFFIC SAFETY FROM AN ECONOMIST'S POINT OF VIEW	
STEFAN VALAVANIS	477
ACCELERATED INVESTMENT AS A FORCE IN ECONOMIC DEVELOPMENT.....	
HOWARD S. ELLIS	485
AN ECONOMIC JUSTIFICATION OF PROTECTIONISM	
EVERETT E. HAGEN	496
A STUDY OF DECISION-MAKING WITHIN THE FIRM	
EDWIN MANSFIELD and HAROLD H. WEIN	515
THE PLACE OF MONEY CAPITAL IN THE THEORY OF PRODUCTION.....	
ANDRE GABOR and I. F. PEARCE	537
CONTROLLING CONSUMERS IN THE EVENT OF FUTURE WARS	
STEPHEN ENKE	558
COMMERCIAL POLICY AND ECONOMIC NATIONALISM	
FRANK H. GOLAY	574
UNITY IN THE THOUGHT OF ALFRED MARSHALL....	
G. PURSELL	588
NOTES AND DISCUSSIONS	
Culbertson on Interest Structure:	
Comment.....	L. S. Wehrle 601
Reply.....	J. M. Culbertson 607
The Effect of Devaluation on the Price Level:	
Comment.....	Jaroslav Vanek 614
Further Comment.....	Randall Hinshaw 616
Reply.....	Egon Sohmen 626
A Note on Professor Solow's Growth Model.....	John Buttrick 633
Recent Publications.....	637

PUBLISHED BY HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS

The Quarterly Journal of Economics

Published by Harvard University

Editor

ARTHUR SMITHIES

Associate Editors

ABRAM BERGSON
JOHN D. BLACK*
J. K. BUTTERS
T. N. CARVER*
EDWARD H. CHAMBERLIN
ROBERT DORFMAN
JAMES S. DUESENBERY
JOHN T. DUNLOP
JOHN KENNETH GALBRAITH
ALEXANDER GERSCHENKRON
GOTTFRIED HABELER

ALVIN H. HANSEN*
SEYMOUR E. HARRIS
CARL KAYSEN
WASSILY LEONTIEF
EDWARD S. MASON
GUY H. ORCUTT
SUMNER H. SLICHTER
OVERTON H. TAYLOR
ABBOT P. USHER*
JOHN H. WILLIAMS*
EDWIN B. WILSON*

* Emeritus

Assistant Editor

CATHERINE RUGGLES GERRISH

Books, periodicals, and manuscripts to be addressed, THE QUARTERLY JOURNAL OF ECONOMICS, M-12, Littauer Center, Cambridge, Mass.

Business letters to be addressed, HARVARD UNIVERSITY PRESS, 79 Garden Street, Cambridge 38, Mass.

Subscriptions \$5.00 a year. Three years, \$14.00. Single copies (except those published in two parts) \$1.50, postpaid. A special subscription rate of \$3.00 is offered to students. Such special subscriptions may be for one, two, three, or four years. Those desiring to take advantage of this reduced rate should include with their remittance a statement from an officer of their college or university that they are *at that time* "regularly enrolled students who have not yet received the Doctor's degree."

CONTENTS FOR AUGUST, 1958

THE EXTREMITIES OF CURRENT AGRICULTURAL POLICY PROPOSALS.....	John D. Black
THE RATE OF INTEREST IN A DYNAMIC MODEL.....	A. Ll. Wright
THE ANATOMY OF MARKET FAILURE.....	Francis M. Bator
A CRITIQUE OF CONCEPTS OF WORKABLE COMPETITION.....	Stephen H. Sosnick
ECONOMIC STRUCTURE AND THE REGULATION OF TELEVISION.....	Harvey J. Levin
NOTES AND DISCUSSIONS	
Peak Loads and Efficient Pricing:	
Comment.....	Jack Hirshleifer
Further Comment.....	H. S. Houthakker
Reply.....	Peter O. Steiner
Investment Criteria and Capital Intensity Once Again.....	Albert O. Hirschman and Gerald Sirkin
The Rockefeller Report: Implications for Fiscal and Monetary Policy.....	Arthur Smithies

Recent Publications

COPYRIGHT, 1958, BY THE PRESIDENT AND FELLOWS OF HARVARD COLLEGE

Entered as Second-Class Mail Matter. Acceptance for mailing at special postage rate provided for by Section 1103, Act of October 3, 1917, authorized on July 31, 1918

Published by the Harvard University Press, Cambridge, Mass., U.S.A.

London: Oxford University Press

THE QUARTERLY JOURNAL OF ECONOMICS

Vol. LXXII

November, 1958

No. 4

TRAFFIC SAFETY FROM AN ECONOMIST'S POINT OF VIEW

By STEFAN VALAVANIS

I. Economic reasoning applied to traffic safety, 477. — II. Inconsistencies in costing the evils of driving, 478. — III. Marginal cost taxation and efficiency, 479. — IV. Expenditures need not be tailored to revenues, 480. — V. Differential insurance premiums, 482. — VI. Accidents are social overhead costs, 483. — VII. Optimizing our institutions, 484.

I. ECONOMIC REASONING APPLIED TO TRAFFIC SAFETY

This article is stark because I wrote it for noneconomists¹ and because I normally prefer to state my case bare and worry later about the refinements. I do not really believe one hundred per cent of what follows.

The economist's professional reasoning has seven guiding principles to contribute to the problem of traffic safety: (1) He sees it not as an "all-or-nothing" problem, but a "best mixture" problem whose solution involves *some* accidents plus *some* traffic. No accidents, or no traffic are out of the question. (2) He is used to thinking in terms of clever sets of rewards and punishments to make men behave in a socially desirable way, even if they seek only their own selfish ends. He recognizes that rewards and punishments hit diminishing returns and acknowledges that there is room for culture, education, cure, remotivation, etc. (3) He prefers to decentralize decisions rather than plan from the top. This is both good administration and an end in itself in our society. He would rather set prices and let people take things or leave them. (4) He considers prices (taxes, insurance rates, etc.) mainly as rationing devices and only secondarily as means of raising revenue or covering cost. (5) He refuses to opti-

1. Twenty-one representatives of various disciplines from divinity to cybernetics discussed traffic safety and proposed research undertakings at a conference sponsored by the National Safety Council under the auspices of the President's Committee on Traffic Safety, in Williamsburg, Va., Feb. 23-28, 1958. This article is slightly reworked from a memo I circulated at that conference.

mize on the most valuable resource (say human life) as though all others (for instance, time, gasoline, concrete, traffic signals) were free. To him everything has an exchange rate (usually a variable one). (6) He abhors the approximating, maximizing or minimizing of some magic ratio like accidents per passenger-mile. (7) He prefers to maximize imperfectly in the large, not perfectly in the small. Perfecting private car traffic is worthwhile, but other ways of getting transportation, prestige and pleasure besides car-ownership may be more worthwhile, even if imperfectly organized.

II. INCONSISTENCIES IN COSTING THE EVILS OF DRIVING

The inconsistencies are of three kinds: those of political action, external effects, and consumer imperfection.

Organized society is inconsistent in costing accidents and other evils of driving. The same individual can be a meek pedestrian and also a fierce driver, or a lenient judge. He demands stern traffic policemen, but does not mind "fixing" the traffic tickets he gets. As a driver, he wants better highways; as a voter, lower taxes. To prevent a death from polio, society spends millions; the same society often refuses to install a traffic light at a dangerous corner. The cost of a death is much higher to the family than the amount of the judge's award, and to society than the cost of replacing the deceased. When the driver happens to be insured, often the court, garage and hospital pad indemnities, repair charges, and medical bills. Damages are sometimes only compensatory, and at other times, in similar circumstances, sharply punitive.

Social and private evils cannot all be minimized together: death, permanent injury, jail sentences, slow traffic, policemen's wages, exhortations, driver education, etc., are evils, substitutable to some extent. Most people agree to a hierarchy of evils, but problems can be tackled even if they do not.

External effects arise in traffic situations: (1) When there is harm to others who cannot be compensated, as when one's mere presence helps congest the streets. (2) When others benefit at the same time the individual does, but they cannot be charged as when one sweeps a pavement or waters a garden. This is why yielding the right of way is a matter of law, not of whim. (3) When the consumer wants to harm himself only, but society hates to see him do it. This is why we ban the game of "chicken" even if the contestants are without family, have prepaid their funerals, and selected a perfectly empty road to play on.

Examples of consumer imperfection: (1) People allegedly under-

estimate the risk of jaywalking. We ban jaywalking for the same reason that we require a doctor's prescription for dispensing a drug. (2) People are sometimes under temporary disability; for instance, the driver may be well aware of all the risks but is in a hurry.

With external effects and consumer imperfection we feel justified in denying the consumer his sovereignty (e.g., taxing whiskey) though possibly leaving him free choice (to buy or not to buy the taxed whiskey).

III. MARGINAL COST TAXATION AND EFFICIENCY

Prices, including insurance premiums, taxes, fines, etc., are primarily not to cover costs or to raise revenue or to distribute income, but *to ration things and to show where the need is greatest*. From the point of view of rationing, the price of mineral water is justifiably high even though its cost is zero; once a bridge is built it is better to allow free use of it, rather than have it partly idle; meter parking fees are to discourage parking, not to pay the cost of meters; prices do not have to cover the sum total of costs. Such principles ought to govern insurance rates, taxes, license fees, stop signs and other impediments to motor traffic.

The fundamental proposition of welfare economics says: "For maximum efficiency in the short run, the price of everything should be its marginal cost (*not* average cost plus markup). For maximum efficiency in the long run, invest in the directions where marginal cost exceeds average cost the most." The following example (drastically simplified) illustrates these principles.

Suppose we agree that the death of any one man costs (in some sense) \$100,000 and the probability of such an accident increases² with the number of cars as shown in the table on the next page.

Under the marginal cost pricing rule, if all cars are alike, each should pay \$100. Then aggregate insurance premiums are $\$100 \times 1001 = \$100,100$, while aggregate claims paid to victims are $\$100,000 \times 1.1 \text{ per cent} = \$1,100$, and there is therefore an excess collected of \$99,000.

Any other method of pricing is uneconomical, because each prospective 1001-st driver would pay less (or more) than the damage the court says he has occasioned.

The argument illustrated by the example is perfectly general: The more cars (or car miles, or passenger miles) the greater the likelihood of death. Therefore marginal cost is higher than average cost.

2. With still more cars on a fixed roadbed system, traffic slows down, possibly reducing the probability of death, while costs of delay and frustration go up.

Our conventional insurance premiums are based on average cost plus a small markup. *Conclusion:* Our insurance rates are much too low and hence inefficient. It is as though our insurance commissioners

	Per Cent	Total Cost	Average Cost per Car	Marginal Cost
With 1,000 cars on the roads the probability of one death is	1	\$1,000	\$1.00	
With 1,001 cars on the roads the probability of one death is	1.1	\$1,100	\$1.10	\$100
(No other factors involved. Highway net, car models, etc., given.)				

contradicted what our judges said a man's life was worth. What has been said about the 1001-st driver can be said about the 1001-st pedestrian: his presence raises marginal cost — but he pays nothing. We do not make the pedestrian pay his marginal costs in cash for several reasons: (1) these costs are likely to be small; (2) they are difficult to administer; (3) we put minor nonmonetary vexations in the pedestrian's path — he has to wait for traffic lights to change, he gets elbowed, splashed with mud, etc. These latter factors determine (other things being equal) whether he will walk or drive.

IV. EXPENDITURES NEED NOT BE TAILORED TO REVENUES

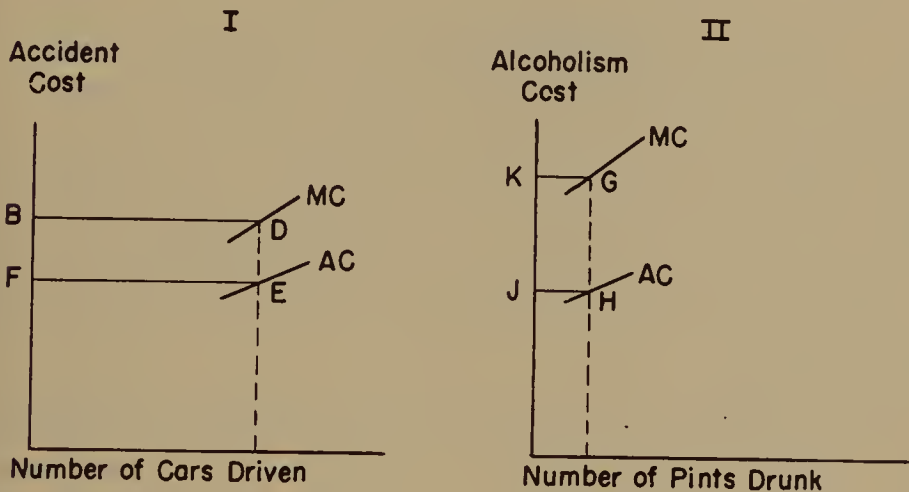
Now consider that mysterious excess of \$99,000. How are we to dispose of it? Of course, it is not to be kept by the insurance company. If insurance companies are allowed a given mark-up to cover administration and moderate profit, they will have an interest in seeing the average cost reduced, because they pocket the difference between the premium and the lowered average cost and there is always a lag in adjusting premiums to realistic average costs. The excess of \$99,000, like the revenue from any tax, is to be spent for some public purpose. But what?

I was tempted to exclaim: "How nice of economic theory to provide us with a fund from which to educate, research, motivate, drivers and pedestrians, and to invest in better highways! Moreover, we do not have to worry about how much to spend. The answer (\$99,000) is inherent in the problem. Were not our statesmen right when they said: 'Use liquor tax receipts to combat alcoholism, gasoline tax revenues for highway improvement, and in general link the tax to the use, the revenue to the benefit?'"

As a matter of fact, \$99,000 is almost certainly the wrong

amount to spend, because it is highly unlikely that out of the total funds spent on this and that public purpose \$99,000 is proportional to accidents' relative gap of *MC* over *AC*.

In the figure, let I represent insurance pricing and II liquor taxation: and let there be (for simplicity) no other public problem but traffic and alcoholism. Vertical axes measure the cost of traffic and alcoholism respectively. Horizontal axes measure intensity of



traffic and drinking. The gap *GH* of marginal over average cost is so much larger in alcoholism than the corresponding gap *DE* in accidents, that we probably ought to be spending not only the whole liquor tax revenue, *JKGH*, but perhaps also a large part of the motor car insurance "tax" revenue (*FBDE* or \$99,000) on the cure and prevention of alcoholism. It is logically and ethically absurd to devote to the cure of an evil only such funds as can be collected in taxes from those most enslaved to it, i.e., people who would rather pay the tax than go without alcohol. The rule is: spend in the direction where the gap is large, not where revenues are large. Likewise it can be argued plausibly that cigarette taxes should be spent on highways, income taxes and fines on the national ballet and so forth, according to the size of discrepancies³ in marginal over average costs (and benefits) in these various directions. The entire budget ought

3. Some economists go even farther: if *MC* is less than *AC*, it pays to subsidize the activity. For instance, if it is true that the *AC* of treating alcoholics decreases with their number — because better use is made of nurses, doctors, psychiatrists, etc. — then tax traffic (a greater evil) and subsidize the price of liquor (a lesser one). Indulging in alcoholism when specialized nurses are idling is more economic than trying to use the highways when they are already choked with traffic.

to be interdependently optimized, not split into compartments by source of revenue.

V. DIFFERENTIAL INSURANCE PREMIUMS

Ideally the tax on driving should be by area and time driven, being higher in the congested areas and rush hours; less ideally, one could also charge a graduated tax on annual mileage.

Most of our present insurance cost differentials are uneconomic. It is not easy to justify differentials by *residence* area, so that a Boston driver pays, say, twice as much as a resident of Concord or Lowell. Unless cars remain in closed nonoverlapping regions, all cars contribute to Boston's congestion and fatalities. An extra death is equally expensive no matter who contributes to its probability of occurring. All cars are culprits (to the extent that their travels overlap) whether the owner resides in, or commutes to, Boston.

Differentials can be very efficient, moderately efficient, or not at all efficient. They can be equitable or inequitable independently of their efficiency.

For example: a tax on driving in sleet is highly efficient, because sleet is more hazardous, and roads should be thinned of traffic in favor of those whose business is so important as to warrant paying the tax — most people do not have to drive in sleet, and they can postpone a relatively unimportant trip — i.e., most people can choose to escape the tax by shifting over to the untaxed class of drivers. Let doctors also pay, and let the price of medical attention in bad weather reflect the tax; for a hypochondriac or a patient with a minor complaint does not rate a cheap doctor's call during sleet. Second example: taxing the very young, the very old, women, etc., more heavily than others is efficient, but less so, because though it eliminates all but the most eager of the youthful, senile, or female drivers, these people cannot by an act of choice change their age and sex to escape the tax, and incidentally lower cost. Similarly, tax the poor worker driving to work during rush hour, and let his wage, and the price of his product reflect the fact that the plant wants to operate between rush hours. Not to tax the worker is tantamount to subsidizing the price of goods that otherwise have no business being produced between 8:00 A.M. and 4:00 P.M., rather than between 7:45 A.M. and 3:45 P.M. If the worker cannot shift the tax to the price of the product and we feel that he is unduly burdened, this is an argument in favor of income equality, not an argument against marginal cost taxation of traffic!

The rule for equity in differential insurance rates is: apply them

where the insured can (within reason) remove himself from a high to a low risk category, by getting a better car, driving to work ten minutes before or after rush hours, staying indoors during storms, etc.

If I belong to a low risk category, it does not mean I shall not have an accident, for in this world, anything can happen. It is a mistake to penalize excellent risk drivers just because they have had an accident. One must penalize the *probability* of an accident, not its *having happened*.

VI. ACCIDENTS ARE SOCIAL OVERHEAD COSTS

After ideal research has identified the big factors that affect the probability of a traffic accident, millions of little, independent factors remain to constitute the "pure accident," a random residual. How to finance this hard core, which may be quite substantial? One answer is, out of the "excess" of marginal over average cost of the identified factors. But the excess may not be enough, because it relates to the assignable costs (one identifiable factor at a time) and the residual may be very large. For instance, if two systematic independent measurable positive numbers, factors f_1 and f_2 ($0 \leq f_1, f_2$) are at play, plus a random one, r (or residual factor), the probability p may be

$$p = 1 - e^{-(f_1 + f_2 + r)}$$

and the variance of r may be immensely larger than the variance of either of the two identified factors. Then the total charges as prescribed by marginal pricing

$$f_1 \frac{\partial p}{\partial f_1} + f_2 \frac{\partial p}{\partial f_2}$$

may not be sufficient to cover the costs occasioned by f_1 , f_2 , and r .

Moreover, even if they were sufficient, benefits from driving accrue not only to the motorist, but to society at large in the form of more travel, a more flexible labor market, less neighborhood parochialism, better civil defense evacuation and so forth. Then car accidents are the price society must pay for progress, etc., and so why penalize the driver? Accidents are the other side of the coin of progress, just as are drunks, the insane and neurotics that result from our complex civilization which necessarily breeds some misfits, given man's limitations. Far from condemning such people for their failure, we support them from the public purse. Likewise, a case can be made for meeting the hard core of traffic costs from general

taxation, rather than by assessing the residually unlucky driver, or by abandoning his victims to their fate.

VII. OPTIMIZING OUR INSTITUTIONS

What I have advocated so far is within present institutional constraints, for example, that drivers shall not be required to take strange passengers; that they may drive anywhere, at any time if willing to pay, and so on. True, a practical man would take the motor car and the North American's private attitude towards it as given and try to do the best under this constraint. But again, why take them as given? What if it can be shown that a better solution is available outside our current institutional confines?

For instance, a case can be made for roping off to private wheel city traffic centers, national parks, summer resorts. We have evidence that our tastes for cars interact positively: if so, there is a case for teaching ourselves all around not to like them so much, and to get the same level of satisfaction with fewer, smaller, less powerful ones. Such reforms are less fantastic than the car's recent transformation of the face and attitudes of North America. And after all, the age of the motor car, like that of the dinosaurs, is probably a queer interlude, a mere exception. The Soviet world is likely to skip it altogether and proceed directly to helicopters; or not even to them. Why cannot we rearrange cities and places of work so as to save 90 per cent of all mileage, which is to and from work? Such better solutions, even if they exist, cannot be reached if we merely try to improve and perfect our suboptimal system.

Public concern, action committees, investigations are often evidence that someone is getting shortchanged and that redress is possible only by political and missionary action, not by market haggling. Traffic problems have reached this stage.

HARVARD UNIVERSITY

Stefan Valavanis was shot by an army deserter in Greece on July 18, 1958. His charm, his wit, and his originality, combined with the tough fiber of his character made a lasting impression on everyone who knew him. His death at the age of thirty has cut short a career of brilliant promise.

ACCELERATED INVESTMENT AS A FORCE IN ECONOMIC DEVELOPMENT*

By HOWARD S. ELLIS

Introduction, 485. — I. The demographic argument, 485. — II. The role of the propensity to consume and savings, 488. — III. The importance of external economies, 491. — IV. Overtones for policy of the doctrine of accelerated rates of investment, 492. — V. Conclusion, 493:

I

An increase in investment characterizes all economic development: indeed, aside from the rise of income per capita, it is probably the most central phenomenon. At first glance, therefore, it might be somewhat puzzling why theories which stress this factor should be in any way distinctive as theories of economic development.

In part the answer lies in relative emphasis. From the body of classical economics, one gathers the general impression that the main-spring of progress is the slow but steady accretion of capital through individual saving, i.e., parsimony or frugality. Marx laid the primary emphasis upon the "mode of production," the organization of production; and Schumpeter, Spiethof, and many subsequent writers have thought the innovating entrepreneur to be the *primum mobile* of economic growth or development. With Kuznets, who is indeed very loath to commit himself regarding causation, there seems to be a good deal of emphasis upon the increase of population, either internally or through immigration, and the possibility that it is a main independent variable. Other economists have stressed the gains in productivity, technical knowledge, and motivation springing from international trade and finance. And one must not fail to take account of theories which discover the most important impulses in noneconomic factors — in political or social revolutions, ideologies, and the like. But there is a congeries of theories which — while not necessarily ignoring or denying the operation of one or all of these forces — lays the weight upon the *speed* of the investment process; indeed, in some cases this emphasis is almost exclusive. If capital formation is rapid, other attributes of economic progress are practically certain to be realized; if capital formation is not *rapid*, economic development will be slight or fail to materialize altogether.

* The author is obliged to the editors of *Revista Brasileira de Economica* for permission to publish in English an article originally appearing in Portuguese.

But, in addition to this emphasis, the theories I am considering have several characteristics which warrant their being grouped together. For one thing, these theories are generally strongly interventionist, at least so far as concerns the assumption of responsibility by the state for a greatly increased rate of saving, and — extending out from this basis according to the predilections of the individual writer — to more or less, and generally more, control (and sometimes operation) of the specific lines of investment and production. Secondly, these “big push” theorists usually consider manufacture as inherently superior to primary production as a vehicle of development. These two characteristics are so general that I shall terminate the list with these alone for greater emphasis; but it would be tempting to point to the frequency also of an inflationary bias in writings of this sort, to autarkical leanings and to a fondness for general equilibrium planning as implied by linear or nonlinear programming. But the interventionist and other features of these theories, upon which I shall want to comment later, are their overtones rather than their substance.

The substantive bases for an accelerated rate of investment through state intervention are principally three: a demographic argument, a line of reasoning involving the propensity to consume (or to save), and thirdly, conclusions reached from the technical discontinuities or “lumpiness” of investment. Let me say clearly in advance that in no case do I reject the reasoning completely; but that in all cases I attach much greater weight than do the proponents of these theories to the limits of possible gain, to the risks and costs of the proposed line of action, and to the merits of alternative policies.

I turn first to the demographic argument for the big investment push, because here both the merits of the theory and its detractions are most conspicuous. The setting of the problem is an economy in which per capita incomes are low, capital is scarce and labor abundant, death rates have recently been lowered by modern methods of public health, and birth rates are high — a milieu highly inauspicious for economic development. In such a setting, the case for the *investment* attack has been most ably argued by Professors Galenson and Leibenstein.¹ According to these writers, the only hope of eventually raising per capita incomes is temporarily to abandon the usual marginal productivity criterion for investment and in its place to

1. Walter Galenson and Harvey Leibenstein, “Investment Criteria, Productivity, and Economic Development,” this *Journal*, LXIX (Aug. 1955), 343–70; see also Harvey Leibenstein, *Economic Backwardness and Economic Growth* (New York: Wiley, 1957), chap. 8.

maximize the "marginal per capita reinvestment quotient," i.e., adopt techniques which "lead ultimately to the maximum capital/labor ratio." The reinvestment quotient should be maximized even if there is surplus labor in the economy because, by definition, techniques of the sort appropriate to development are those which lead most speedily to the maximum of capital equipment per laborer and hence to maximum real wage rates.

One objection to this theory of development is that the ratio of capital to labor can *always* be raised and, short of some very remote limit, real wage rates can *always* be raised by more capital, so that the term for the application of the reinvestment quotient is indefinite. A second objection is that the theory ruthlessly sacrifices present gains in productive capacity to the furtherance of productive capacity. Galenson and Leibenstein are quite frank about this; indeed, it is partly their fear that any increase of *present* real wages will lead only to an increase of the birthrate that leads them to propose this policy. They are willing to accept even an increase of unemployment for the sake of techniques maximizing the ultimate ratio of capital to labor.

Now it is indeed quite possible that concern with unemployment may be excessive and may interfere with necessary reforms in agriculture and in industrial techniques.² "In theory," at least, additional unemployment caused by technological improvements can be absorbed, just as with the basic substratum of structural unemployment, by public works of a local, labor-intensive variety. But there are limits to the speed with which these projects can be planned and executed, and there are corresponding limits to the desirability of labor-saving techniques. Indeed, the employment of labor-saving, capital-intensive techniques in an economy with surplus or even cheap labor is an anomaly. The anomaly is basically that the government, in introducing these techniques, is discounting the future at a lower rate than the market.

I am not prepared to argue that this is never appropriate; indeed I would myself maintain that a good deal of the initiative for saving has to be taken by governments in the early stages of development through raising funds by taxation. But there are limits, outside the harshest totalitarian states, to this forcing of saving, first upon ordinary ethical considerations of human welfare and freedom, and secondly upon the pragmatic basis that the population may simply not tolerate an indefinite postponement of the fruits of progress. Thus the objection to Galenson's and Leibenstein's argument is partly a

2. Ragnar Nurkse, "Reflections upon India's Development Plan," this *Journal*, LXXXI (May 1957), 188-204.

matter of degree; but it is also partly a matter of their cloaking under a phrase, "the marginal reinvestment quotient," which has the air of an objective investment criterion, the purely arbitrary decision of authority to plough back productivity gains into plant and equipment for whatever period may be pleasing to that authority.

If their concern lest population increases absorb the gains of progress leads these authors to an intolerably stiff policy, the very opposite is probably the shortcoming of the doctrine proposed by John Stuart Mill and Frank W. Taussig. There are grains of a "big push" thesis in this demographic theory also. If productivity could be thrust forward rapidly, it was argued, by whatever means, the immediate impact in raising real wages would give the population a taste for greater material comfort, which they would defend by reducing the size of their families. The trouble with this prescription, which might have worked and probably did work in times past, as, for example, in the industrialization of Japan, is that in the current scene the increase of population from lowered mortality rates is apt to eat up productivity gains before they can be transmitted into higher *levels* of living. This leaves the brunt of responsibility upon a rise of the *standard* of living, which, as Arthur Lewis has recently insisted, is a slow process.³ But it is at least a process which has worked in the history of Western civilization, and has worked without compulsion upon the mass of humanity by the state.

Indeed, upon the assumption that a "big push" of investment means only a temporary resort to compulsory limitation of consumption, what is ultimately to determine birthrates? Will the increase of productive apparatus, such as Soviet Russia is now achieving through "starving the present," result ultimately in higher wages or simply in more people? The answer lies, of course, in the ideals of the population regarding the family, monogamy, continence, and legitimacy, in literacy, sex education, the status of women and ethical and religious beliefs — in a vast complex of forces. Against these basic determinants, a period of accelerated investment for the sake of its demographic effects seems relatively short-lived and superficial.

II

Government-accelerated investment is sometimes justified as the key to economic development on the grounds of the attending behavior of savings. I distinguish four types of this sort of theorizing; but the first, according to which the state may speed up development by enforcing a reinvestment of income continuously into capital goods,

3. W. Arthur Lewis, *The Theory of Economic Growth* (London, 1955), p. 315.

has already been considered sufficiently. This may be called a policy of explicit or openly avowed forced saving.

The second variety relies upon the concealed forced saving of inflation; and I need scarcely remark upon the large number of acquiescent friends or active advocates, even among the ranks of professional economists, that inflation seems to command. As a representative of this school of thinking one may take Professor W. Arthur Lewis of Manchester University, who holds that governments must choose between taxation and inflation in order to provide a necessary supplement to private voluntary saving, and that this choice is "largely political."⁴ Inflation devoted to pure consumption or to the destruction of goods (e.g., war and armaments) become cumulatively worse by reducing the supply of goods relatively to money. "Inflation for the purpose of creating useful capital are on the contrary self-destructive, since sooner or later they result in an increased supply of goods to the market" (p. 217; cf. also p. 405).

Both parts of this proposition are completely indefensible. In the first place "the *purpose* of creating useful capital goods" is generally poorly served by a "big push" of state investment which exceeds tax income plus bond sales to private savers, i.e., by inflation. Inflation, as Schumpeter so eloquently sets forth in *The Theory of Economic Development* biases the economic calculus by making estimates fallible and by making profitable, however well-motivated the original investments, all sorts of speculative activities through the multiplier or velocity-of-money effect. Lewis has himself dwelt upon some of these costs of inflation, including balance-of-payments deficits and the accompanying necessity of exchange controls, etc. The increased flow of goods *need* not be, and in all probability *will* not be, sufficient to prevent a continuous rise of prices as long as inflation "for the purpose of creating useful capital goods" is pursued.

Indeed, the phrase "sooner or later" applied by Lewis to the "increased supply of goods to the market" reveals that, even aside from the various costs of inflation in inefficiency and the reduction of voluntary saving, inflationary financing of a given increment to investment is only noncumulative if it is a "once-over" performance. The yield of investment comes in slowly over the ensuing months or years and does indeed, if the investment is well conceived, give rise to an amount of deflation sufficient to offset the original investment (assuming with Lewis that it has been financed by inflationary means). But a repetition of inflationary increments to investment clearly produces cumulative inflation. I therefore take leave of the idea that a "big push" automatically generates its own saving.

4. W. Arthur Lewis, *op. cit.*, pp. 217 *et seq.*

Quite a different line of thought is represented by those who believe that consumption lags in point of time behind variations of income because consumption is somewhat a matter of habit. Thus if income can be given a sudden forward thrust by a "crash" investment program, voluntary savings will for a while be a larger fraction of income, and will thus lay the ground for further advance.⁵

It is amusing to observe that the assumption that consumers do not keep pace with the rise of income is the direct opposite of the assumption made by those who argue that a sudden increase of income will, by raising the level of living, instill higher standards of living and thus reduce the birthrate. Both arguments for a big investment effort cannot be simultaneously valid. In view of the force of the "demonstration effect" in the modern world and the tendency of welfare legislation and wage demands to equal or outstrip increases of productivity, I should not be hopeful of the magnitude of savings to be achieved through a lag of consumption. But I should also be skeptical of the lowering of birthrates merely through a sudden increase of income, if other social and cultural factors remain unchanged.

The most sophisticated argument, linking an accelerated investment program to increased savings, has been formulated by Gerald Alter, a member of the staff of the International Bank for Reconstruction and Development.⁶ As the title of Alter's paper indicates, he is concerned with the capacity of a borrowing country to service its foreign debt. The model, constructed upon eight variables, shows "that the likelihood of being able to service a larger volume of foreign capital is greater than the likelihood of servicing a smaller volume of capital inflow," a result which "follows from the fact that very small increases in the required marginal savings ratio are associated with very large changes in *per capita* income." Alter emphasizes the "highly favorable savings effect which a 'big push' makes possible"; and there seems to be no question that his model could be extended from the servicing of foreign debt to internally held public or private debt.

But the author himself clearly points out that the favorable outcome of accelerated investment is a possibility and not a certainty.

5. This seems to be the type of discontinuity with respect to savings which Paul Rosenstein-Rodan has in mind; cf. his "Notes on the Theory of the 'Big Push' " (mimeographed), Roundtable of the International Economic Association, Rio de Janeiro, August 1957.

6. Gerald M. Alter, "The Servicing of Foreign Capital Inflows by Underdeveloped Countries" (mimeographed), Roundtable of the International Economic Association, Rio de Janeiro, August 1957.

Less favorable behavior in the basic variables could reverse the outcome, as, for example, if the capital output ratio should rise, the required amortization period for the debt should be shortened, population growth accelerated, or rates of interest on borrowed funds increased. The chance for achieving greater income by expanded investment is matched by greater risks. Thus Alter recognizes one of the "facts of life" all too frequently ignored by proponents of massive investment programs. Risk is a genuine economic cost, and its rational appraisal might easily account for the rejection of large-scale ventures in either the private or public spheres. Too great sensitivity to risk could indeed block progress; but this truth scarcely warrants the conclusion of a recent writer that "Extra funds for investment should then be fearlessly extracted from the consumer and splashed about."⁷

III

The chief basis upon which the "big push" of investment has been justified, since its original enunciation by Paul Rosenstein-Rodan a decade and a half ago,⁸ has been the possibility of realizing extensive external economies, and this ground is still a favorite with nearly all writers of this persuasion. But the great offset to the possibility that domestic development programs should give rise to further external economies has been definitively set forth by Professor Viner: foreign trade makes available to the developing country the much more substantial economies realized upon world markets, independently of home investment.⁹ This fact is now recognized by Professor Rosenstein.¹ But he fails to give overt recognition to the further fact adduced by Viner that the newly developing countries nowadays are chiefly primary producers, and, as such, investment for exports and for marginal import substitutes, where external economies are presumably negligible, occupies a very large part of total investment. For this entire sector, the "big push" loses its specific justification from external economies.

We are left then with that portion of production for the domestic market which does not substitute for imports. Still, this can be a very substantial field, embracing purely local consumer goods pro-

7. P. Wiles, "Growth *versus* Choice," *Economic Journal*, LXVI (June 1956), 254.

8. P. N. Rosenstein-Rodan, "Problems of Industrialization of Eastern and South-eastern Europe," *Economic Journal*, LIII (June-Sept. 1943).

9. Jacob Viner, "Stability and Progress: the Poorer Countries' Problem," First Congress of the International Economic Association, Rome, September 6-11, 1956; mimeographed paper, pp. 27-31.

1. Rosenstein-Rodan, *op. cit.*

duction and most public utilities — transportation, communication, power, water and sewerage facilities, and the like. Even here, however, there are limits to potential external economies. Viner points out that certain investments — presumably in the case of fairly inelastic demand — are cost-reducing rather than output-expanding. Since external economies depend upon expansion of output in the initial industry, they become negligible for this category of investment. I should like to call attention to two further limitations of considerable significance. In the field of purely domestic goods, a large fraction will be personal services and very light industry (a good deal of food and raiment production) in which the “chunkiness” of fixed investment is unimportant because fixed investment is itself a small fraction of costs. Since external economies are simply internal economies in adjacent industries, their significance is correspondingly small in these cases. It is furthermore worth remembering that, in the case of public utilities, potential external economies do not pertain to the cost of the *equipment* of these industries if it can be more cheaply imported.

Taken together, all of these limitations need not entirely remove the possibility of external economies. But they are neither as universal as often supposed nor, when they actually exist, as substantial. Furthermore and finally, though their existence does increase the productivity of the economy for given magnitudes of investment, they do not constitute a reason for a *concentration of investment in point of time* if — as would appear probable in any but the smallest countries — the “chunkiness” of individual investments levels out to a fairly full utilization of capacities in the aggregate for all capital facilities together. This is a decidedly relevant consideration if “accelerated investment” is taken, not as simply synonymous with more investment continuously, but as a “big push” followed by a lower rate.

IV

Beyond its substantive theoretical basis in the population, savings, and external economies arguments, the doctrine of accelerated rates of investment has overtones for policy which its proponents, I am sure, would not be content to have ignored. One of these is the predilection for manufacturing over agricultural and other primary industries. In part this predilection may simply reflect a sentimental desire to see the country “independent” of its neighbors, particularly the richer ones; but in part it may rest on rational arguments, such as the improvement in labor morale which is supposed to attend factory

production, the cultural and demographic effects of large cities, which are supposed to be favorable to economic progress, and the risks of primary production from the fluctuations of world markets. On the other hand, agricultural and primary types of production have in their favor that they utilize the relatively abundant factors of land and labor and economize capital; that characteristically in the less developed countries they provide two-thirds or more of the national income; and that, by the same token, they supply the chief where-withal for industrial imports and investment in general.

It would scarcely seem necessary at the present stage of the debate concerning economic development to say that the merits of investment in agriculture versus industry have to be settled according to the peculiarities of each country. By consequence, whatever merits may inhere in crash programs of investment may just as well be associated with agriculture — irrigation, drainage, transportation facilities, reform of fragmented land-holdings, etc., — as with building industrial plants; in particular cases, indeed, more so.

Somewhat similar reflections would be germane to the penchant of the “big push” economists for planning, state direction of investment, and extensive controls. Linear programming, for example, is essentially an information service, and the benefits of its information may just as well be made available to private as to public entrepreneurs. In and of itself, linear programming does not supply any rationale for accelerated investment. If it should appear desirable to supplement private voluntary savings by the fiscal arm of the state, the funds can be lent to private firms. The theoretical underpinning of accelerated investment programs pertains to a *rate* of investment, and not necessarily to government controlled investment. Ordinary economic motivations of the individual and the firm are a powerful engine of economic progress. It would be regrettable if the economists of the free world created an impression to the contrary.

V

What, in conclusion, may be said of the general merits of the “big push” philosophy of economic development? As a starting point for development some kind of impulse is, of course, necessary; a change from stagnation is not likely to come by almost imperceptible degrees. Economic historians and cultural anthropologists have pointed to various prime movers in economic change: to the roles of the foreign trader and foreign capital, to immigration and the transfer of techniques, to the process of technical innovation itself, to cultural change, and to political revolution. Among these, intensive

programs of state investment, as in the Japanese and Russian cases, should certainly take their place. But they are by no means the only or even the chief channel through which development can be achieved; and the demographic advantages, the capital accumulation, and the external economies to be expected from crash programs of government investment can easily be overrated.

A statistical summary of recent economic development throughout the world by John H. Adler reaches the important conclusion, among others, that "a relatively low level of investment 'pays off' well in the form of additional output."² The author emphasizes this conclusion most sharply in connection with India and Pakistan; but the chief reason for this conclusion, the prevailingly low capital-output ratio, is also characteristic of many other of the less developed countries of Asia and Latin America as his statistics reveal. Thus it appears that it is far from generally true that a massive injection of capital is a precondition of growth.

A general weakness of the "big push" doctrine is that it frequently ignores the conditions for *evoking* the investment to which it ascribes such potency in the general picture of development, as well as neglecting the conditions under which investments, once made, can be fruitful. It is through the assumption of a *deus ex machina*, the state, which does all or most of the investing, that this theory is able to avoid the problems of securing not only the saving, but also the willingness to undergo risk, which is implied in investment. And it is only through a singular narrowness that the theory often implies that it tells the whole story of the successful operation of the economy, once the investment is made.

In point of fact, the conditions for the evoking of private investment and the conditions for the profitable use of capital are largely the same. I should place high upon this list the existence of stable and honest government, the absence of inflation, and the accessibility of the economy to the gains of foreign trade and commerce. But other factors, such as the improvement of general and technical education, the amelioration of agriculture (which bulks large in nearly all low-income countries), and progress along the family-limitation front would seem to be equally critical. Taken together, or in some cases even singly, we would seem to have identified a number of factors in economic progress which could outweigh a burst of state-engineered investment.

Some food for thought concerning programs of intensive invest-

2. John H. Adler, "World Economic Growth — Retrospect and Prospects," *Review of Economics and Statistics*, XXXVIII (Aug. 1956), 279; cf. also, 283.

ment would seem to be offered by certain points made recently by Simon Kuznets. His statistical and historical studies lead to the conclusion that "current international differences in *per capita* income are congealed effects of past differences in the rate of growth of *per capita* income." How far would it be necessary to go back into the history of the more advanced countries to reach levels comparable to the *per capita* incomes of the currently less developed countries? The answer is that we should have to go back about ten decades to reach the current income level of Latin America and about fifteen decades for that of Africa and Asia.³ Thus, even at a very early stage in the industrialization of Western Europe, per capita incomes were probably as high as in Latin America today and certainly higher than in Asia and Africa. The economic development of the most advanced countries, at least, scarcely seems to be the result of crash programs.

UNIVERSITY OF CALIFORNIA
BERKELEY

3. Simon Kuznets, "Quantitative Aspects of the Economic Growth of Nations," in *Economic Development and Cultural Change*, Vol. 5 (Oct. 1956); see especially pp. 23-25.

AN ECONOMIC JUSTIFICATION OF PROTECTIONISM*

By EVERETT E. HAGEN

I. Introduction, 496; Manoilescu, Viner, and Haberler, 496; the argument of this paper, 497. — II. The empirical evidence, 498; economic growth and wage differentials, 498; the data, 499; are the differences real? 503. — III. The analytical model, 504; the one-factor case, 504; the two-factor case, 505; extensions and qualifications, 511. — IV. Empirical corollary; protectionism and growth, 513.

I. INTRODUCTION

Manoilescu, Viner, and Haberler

In a 1932 review of Manoilescu's *Theory of Protection and International Trade*,¹ Professor Viner noted that protection of manufactured products would increase real income if a country had "comparative labor advantage" in them but monopoly held wages so high in manufacturing that imports could undersell them.² He added that free trade, which would force the monopolists to reduce their wages and cause the comparative advantage of manufacturing to be revealed in market prices, would accomplish the same effect.

Eighteen years later Viner returned to the subject in one of a series of lectures at the Brazilian Institute of Economics.³ This time he suggested, as appropriate remedies for the wage or price aberration, increasing the mobility of labor in agriculture by providing information and training, and breaking the monopolies, if any, by which

* The thesis presented here was originally presented in a brief note. Discussion of that thesis with a number of persons indicated the desirability of stating the proofs of several steps initially treated as axiomatic or self-evident. These persons include Werner Baer, Francis M. Bator, Richard S. Eckaus, Gottfried Haberler, and Stephen H. Hymer. I am especially grateful to Mr. Hymer for calling my attention to an error in my initial geometry, to Professor Haberler for helpful comments on a draft of the present manuscript, and to Albert O. Hirschman for suggesting elaboration of an ambiguous point. Of course, none of these persons is responsible for errors which may remain.

1. Mihail Manoilescu, *The Theory of Protection and International Trade* (London, 1931). A French edition had appeared in 1929.

2. In *The Journal of Political Economy*, XL (Feb. 1932), reprinted in *International Economics: Studies by Jacob Viner* (Glencoe, Ill., The Free Press, 1951), the argument stated by Viner is not Manoilescu's. Though Manoilescu did not realize it, his basic (and erroneous) equation (p. 103 of his book) reduces to the statement that a country benefits by exchanging agricultural for manufactured products only if its agriculture is more efficient, in some absolute sense, than its manufacturing.

3. Published under the title, *International Trade and Economic Development* (Glencoe, Ill.: The Free Press, 1952).

manufacturing exploits agriculture. Simultaneously, the June 1950 issue of the *Economic Journal* appeared carrying a notable article in which Haberler dealt with the general topic, and three months later Haberler discussed it again in a paper presented to a Round Table held by the International Economics Association.⁴

In his *Economic Journal* article, extending the production-possibility-curve geometry made familiar in the international trade context by the Stolper-Samuelson analysis,⁵ he noted that the test of comparative advantage is not relative money costs in the production of different commodities within a country, compared with relative prices of the commodities if imported, but rather the marginal rate of substitution (marginal transformation ratio) within the country, i.e., the amount of one that can be produced by sacrificing output of the other, compared with the relative price if imported. He assumed intuitively that if wages for equivalent labor (or the unit costs of other inputs) are higher in manufacturing than in agriculture, the exchange ratio (with agricultural products in the numerator) will be greater than the marginal transformation ratio. (The exchange ratio is, of course, the reciprocal of the price ratio.) In this case, welfare in the economy may be increased by producing manufactured products at home, even though they are more expensive relative to agricultural products at home than abroad and manufacturing can survive at home only if protected. He demonstrated that in three cases this may (but will not necessarily) be true. The cases are those of complete or partial factor immobility combined with complete or partial factor price rigidity, external economies, and infant industry. In each case, higher aggregate real income is purchased at the cost of some unemployment.

The argument of this paper

It is the purpose of this paper to generalize and extend the Haberler-Viner argument, and to apply it to the case of economic growth. The argument advanced takes as a point of departure the empirically observed fact that in an economy in which per capita income is rising secularly, the output of manufacturing and mining grows secularly relative to that of agriculture, and the inputs required

4. Gottfried Haberler, "Some Problems in the Pure Theory of International Trade," *Economic Journal*, LX (June 1950), 223-40; and "Real Cost, Money Cost, and Comparative Advantage," *International Social Science Bulletin*, Spring 1951, pp. 54-58.

5. Wolfgang F. Stolper and Paul A. Samuelson, "Protection and Real Wages," *Review of Economic Studies*, IX (Nov. 1941), reprinted in *Readings in the Theory of International Trade* (American Economic Association, 1949).

in manufacturing and mining likewise grow secularly relative to those in agriculture. The same statements may be made of all non-agriculture; for simplicity, a two-sector economy consisting of agriculture and manufacturing will be considered, and a comment about the more complex case will be added at the end of the discussion.

As a result of this secular trend, except in the unreal case of perfect geographic and occupational mobility of labor, wages in manufacturing must be higher than in agriculture. This is true even in the long run, and even assuming complete absence of monopoly in all markets. If they are not higher, manufacturing will not obtain the continuing stream of added labor that it needs. This wage disparity is consistent with full employment. As a result of the wage disparity, manufacturing industry having a real comparative advantage will be undersold by imports when the foreign exchanges are in equilibrium. Protection which permits such industry to exist will increase real income in the economy. However, a subsidy per unit of labor equal to the wage differential will increase real income further, and if combined with free trade will permit attaining an *optimum optimum*.

The proof of these propositions is spelled out in Section III geometrically and where the geometric proof is not obvious algebraically. In Section IV some morals are drawn. Meanwhile, however, empirical background which causes the theory to be of interest is discussed in Section II.

II. THE EMPIRICAL EVIDENCE

Economic growth and wage differentials

Economic growth — continuing rise in per capita income over the long run — is characteristic of the entire Western world and of some Latin American countries, and may now be beginning in China, India, and elsewhere. It is hardly necessary to introduce empirical evidence to demonstrate that in a growing economy wages for equivalent labor are higher in manufacturing than in agriculture, for the logic of the situation is clear. As is well known, with continuing rise in per capita income comes continuing rise in the share of output contributed by industry (manufacturing, mining, construction, etc.) and continuing fall in the share contributed by agriculture. While labor productivity may rise faster in industry than in agriculture, the differential increase in productivity has historically in every country been less than the differential increase in output, and the fraction of the labor force employed in industry continues to rise and that in

agriculture to fall, in the short, intermediate, and long run (except during depressions).

Coupled with this labor force shift is a differentially greater birth rate in rural than in urban areas, by virtue of which the number of laborers who must be pushed out of farm areas and pulled into industry, decade after decade, is even greater than is indicated by the ever-continuing shift in their fractions of total employment.

In this process there exists a force which throughout the short, intermediate, and long run tends to draw the system away from static equilibrium. Only if there were perfect, i.e., instantaneous, geographic and occupational mobility of labor, could returns to labor be equal in agriculture and industry in such a system. Labor is not perfectly mobile. To shift to an urban job new skills must be learned, friends and acquaintances left, and an old way of life abandoned for a new one which, even though it may be equally or more attractive when adapted to, is initially unknown and perhaps forbidding. A wage differential is required to overcome these frictions. And even though labor moves in response to the wage differential, the labor market remains out of equilibrium and the differential must persist. For the industrial demand for labor continues to grow, and new workers must continually be recruited from rural areas. As the supply curve of industrial labor moves to restore the equilibrium wage, the demand curve moves so as to prevent its restoration. Improvement in the knowledge and training of rural workers would lessen the wage differential, but in view of the other impediments to leaving one's home and community to go to a community with different values and customs, such improvement could not eliminate it.⁶

Though the logic of the wage situation seems conclusive, it may be well to present relevant empirical evidence. There is empirical justification for the assumption that real wage rates for equivalent quality labor, even in the long run, are higher in industry than in agriculture.

The data

That money incomes per person engaged are very generally markedly higher in manufacturing, or in "industry," including mining and construction, than in agriculture is abundantly clear. Where data are available in presently underdeveloped countries, they con-

6. Removing monopoly elements that may be present in the determination of industrial wages might lessen but would not eliminate the agricultural-industrial wage differential. Part of the differential often attributed to monopoly may in fact be due to the force sketched in the text above.

Presumably the differential in real wages will depend not only on the degree

firm the fact.⁷ On the basis of national income data for various countries and estimates (some not published) made in the course of research on low income countries, many workers experienced in research in economic development assume as a rule of thumb that income per person engaged in all nonagricultural pursuits combined will in a low income country typically be between 2 and 2.5 times that per person engaged in agriculture including in the latter income in kind, and that income per person engaged in industry, including small scale or cottage industry, will typically be slightly below that in all non-agriculture — say 80 to 90 per cent of the latter.⁸ There are undoubtedly a number of exceptions to this rule of thumb, but it gives a rough idea of the general picture.

Firmer quantitative evidence is presented by Simon Kuznets, in his estimates of the sectorial distribution of income over a long period in fourteen countries,⁹ and in recent years in forty-five countries.¹

of imperfection of knowledge and training, but also on the differential rate of expansion of industry and agriculture. The curve of supply of labor to industry is a rising function of the wage differential, and the curve of demand for labor in industry a falling function of the wage differential. If the rate of expansion of industry rises, the demand curve shifts upward, causing the point of equilibrium wage differential to move upward (greater labor flow) and to the right (higher wage differential) along the supply curve.

7. For example, V. K. R. V. Rao, *The National Income of British India, 1931-32* (London: Macmillan & Co. Ltd., 1940), p. 253, estimates that in 1931-32 income per occupied person was Rs. 124-133 in agriculture, and Rs. 192-195 in industry. Official estimates of the distribution of the Indian labor force in 1951 and of Indian national income by industrial origin for April 1950-March 1951 indicate income per person engaged in industry (including mining and small-scale enterprise) 2.33 times that in agriculture. Since the labor force estimates for industry include all construction whereas the national income estimates may exclude construction activity in commerce, transport, communication and government, the calculation may understate the ratio. The labor force data are from *Papers Relating to the Formulation of the Second Five-Year Plan* (Delhi: Government of India, Planning Commission), p. 237, the national income data from *Estimates of National Income, 1948-49 to 1951-52*, Government of India, Central Statistical Organization, Sept. 1954.

8. I make this statement on the basis of discussion with colleagues at the Center for International Studies and with a few research economists elsewhere. Such bases for estimate as are presented in various volumes written by missions of the International Bank for Reconstruction and Development to underdeveloped countries are consistent with this generalization. The sources concerning India cited in the previous note indicate income per person engaged in all nonagriculture 2.57 times that in agriculture and income per person engaged in industry .90 times that in all nonagriculture.

9. Kuznets presents data for fifteen, but for one of these he has no data for manufacturing separately.

1. For both the recent and long-term data, see his *Quantitative Aspects of the Economic Growth of Nations*; II. *Industrial Distribution of National Product and Labor Force*, being a *Supplement to Economic Development and Cultural Change*, V, No. 4 (July 1957), Appendix Tables 5 and 6.

To conserve space, the data for recent years are not summarized here, except to state that they show a predominance of markedly higher income in manufacturing than in agriculture throughout the world. The fourteen countries for which data extending farther back in time are available are France, Germany, the Netherlands, Norway, Sweden, the United Kingdom, Italy, Hungary, Japan, Canada, the United States, the Union of South Africa, Australia and New Zealand.² In only two of these countries, Australia and New Zealand, where industry did not grow out of agriculture, the two being developed more or less simultaneously by immigrants, do incomes in agriculture show a persistent tendency to be higher than in manufacturing. (This was reversed in Australia during the depression.)³ For eleven of the countries, data are available back to 1900 or earlier. In every one of these eleven except the United States and New Zealand, the ratio of earnings in manufacturing to those in agriculture rose from the nineteenth century to the twentieth. Table I presents simple averages of the earlier and later ratios. The median of the nineteenth century averages is 1.42, of the twentieth century averages, 1.89

TABLE I

TRENDS IN RATIO OF MONEY INCOME PER PERSON ENGAGED IN MANUFACTURING TO THAT PER PERSON ENGAGED IN AGRICULTURE

Country	Early Period		Recent Period	
	Period	Ratio	Period	Ratio
France	1815-1898	1.50	1906-1949	1.89
Germany	1882-1899	1.42	1905-1951	2.16
Sweden	1869-1901	1.81	1909-1951	2.52
United Kingdom	1895	1.08	1911-1954	1.46
Italy	1862-1901	.94	1906-1954	1.63
Hungary	1899-1901	1.66	1911-1943	2.12
Japan	1878-1902	2.29	1903-1942	2.41
Canada	1880-1900	1.23	1910-1953	2.03
United States	1869-1899	2.17	1904-1954	1.67
Australia	1891-1901	.71	1911-1939	1.04
New Zealand	1901	.65	1926-1936	.53

Source: Computed from Kuznets, *loc. cit.*

These figures, of course, reflect profits and other nonwage earnings. The superiority in manufacturing is so great as to make it virtually certain that wage earnings per worker in manufacturing are also generally higher than in agriculture. Even this deduction does

2. In general, the computations for agriculture include unpaid family workers in the agricultural labor force.

3. In Britain, income per person engaged is higher in agriculture after World War II.

not, however, answer the question pertinent here, namely whether wages for jobs of comparable skill — or, for skills quickly learned by workers of a given level of ability and training — are higher than the wages of such workers in agriculture.

With respect to this comparison in underdeveloped countries the opinion of persons who have relevant knowledge of those countries is, to my knowledge, unanimous. Industrial wages in such countries, paid to workers "fresh from the country," are considerably above agricultural incomes.⁴

TABLE II
WEEKLY WAGES, SELECTED COUNTRIES, WORKERS IN AGRICULTURE,
UNSKILLED URBAN WORKERS, AND SKILLED URBAN WORKERS

1 Country	2 Year	3 Unit	4 Agriculture (Male only)	5 Unskilled Urban	6 Skilled Urban	7 5 ÷ 4
Canada	1953	Dollars	37.80	52.86 ¹	66.49 ¹	1.40
Chile	1952	Pesos	633.60 ²	896.52	1220.28	1.41
Denmark	1953	Kroner	132.60 ³	189.12 ⁴	216.96 ⁴	1.43
Finland	1952	Markkaa	3966.00 ⁵	4931.00 ^{1,5}	5562.00 ¹	1.24
Ireland	1953	Pence	1626.00	1542.00	1939.00	0.95
Japan	1950	Yen	1062.00 ⁶	2338.00 ¹	3851.00 ¹	2.20
Portugal	1953	Escudos	121.92	172.80	275.04	1.42
Sweden	1952	Kroner	120.98	190.48 ⁷	226.58 ⁷	1.57
U. S.	1953	Dollars	34.87	70.52	93.93	2.02

Source: International Labour Office, *Yearbook of Labour Statistics*, 1954: For agriculture, Table 19; for urban workers, Table 20. Data for agriculture are for the "complete wage" (worker remunerated wholly in cash). Generally, data in the *Yearbook* for agriculture were for daily wages; these figures were multiplied by 6 to obtain the weekly wage figures shown. For Finland and the United States, monthly wage figures were divided by 4.33; for Sweden, annual figures were divided by 52.

The figures for unskilled labor for each country are a simple average of wages of unskilled workers in those among the following industries for which data were given for the country: textiles, printing and publishing, chemicals, iron and steel, machinery, construction, electric light and power. The figures for skilled workers for each country are a simple average of wages of skilled workers in those among the following occupations for which data were given for the country: loom fixers, cabinet makers, iron and steel melters, machinery assemblers, garage mechanics, carpenters, truck drivers. Except as shown, the figures are for "average wages." For most countries, data were given for several cities. These were averaged. The data for urban workers were for wages per hour. Weekly wages were computed by multiplying by the average number of hours worked per week in the country. Except as indicated, the number of hours used was that shown in Table 15, "Normal Hours of Work in 41 Occupations, October, 1953."

1. Computed by using "actual hours worked per week" in manufacturing (Table 13-A).
2. "Cash including value of board and lodging."
3. "Average earnings."
4. Computed by assuming 48 hours per week.
5. "Minimum rates."
6. Male and female.
7. "Earnings."

The best direct quantitative evidence available consists of wage comparisons derived from the International Labour Office *Yearbook of Labour Statistics*. Table II presents a comparison of wages of

4. W. A. Lewis suggests that wages in the "capitalist sector" are "usually" equal to income in the "subsistence sector" plus "30 per cent or more." See his "Economic Development with Unlimited Supplies of Labour," *Manchester School*, May 1954, p. 150. His observation is presumably based only on casual empiricism, but it accords in general with the empirical observations of many other observers.

agricultural workers and unskilled and skilled urban workers for all countries for which data are available in the 1954 *Yearbook*.

Except for Ireland (where Kuznets' data for recent years show higher income per person engaged in agriculture than in manufacturing), wages of unskilled urban workers are indicated as markedly higher than wages of males in agriculture. Wages for the fairly high-skilled occupations shown are, of course, much higher still. The agriculture-urban differential exists in underdeveloped and economically advanced countries alike; the available evidence suggests that it does not disappear, or even diminish, in the course of development. It is a persistent long-run phenomenon. While the evidence is not absolutely conclusive, the presumption is very strong.

Are the differences real?

One possible explanation of this wage differential of course is that it is not real. Farm wage statistics are difficult of interpretation; the statistics cited may be unrepresentative. Or, the quality of unskilled labor in cities may be higher than that in the country. Or, the excess money income paid to factory workers may be just enough to compensate them for extra money expenses and the lesser conveniences of city life, and for the lesser satisfactions, or greater dissatisfactions, of factory work. Because of the problem of changed tastes after farm workers have shifted to cities, comparison of costs and satisfactions in the two situations is difficult.

But fondness for the good old days and romantic notions of pastoral life should not make us forget that the city has not only inconveniences, but also great attractions for almost any taste, and that the inconveniences (chiefly of transportation) are at their extreme, not at their mode, in the largest cities, which usually come to our minds. It is clear that persons who have experienced both rural and urban life tend to prefer the urban. Few persons who move from farm to city move back; and few persons ever move from city to farm. It is reasonable, therefore, on purely empirical as well as on logical grounds to conclude that the factory-farm difference in money income includes a significant difference in real income. It is plausible to assume a persisting difference in factor payments to identical inputs, or to inputs identical except for negligible training cost.

A model with a wage differential between agriculture and industry, caused only by the dynamics of the system and the drag of imperfect labor mobility, is therefore of great empirical interest.

III. THE ANALYTICAL MODEL

The one-factor case

Conclusions of rather wide importance are drawn here from the assumption of a wage differential. Since they are of a nature that is repugnant to the social welfare "instincts" of many and perhaps most economists, let us consider both a one-factor and two- (many-) factor case, and spell out the proofs with some care.

Figure I illustrates the one-factor case. Let labor be the only input, inelastic in supply. Let A represent the output ("apples") of agriculture, and M the output ("suits") of manufacturing. Let

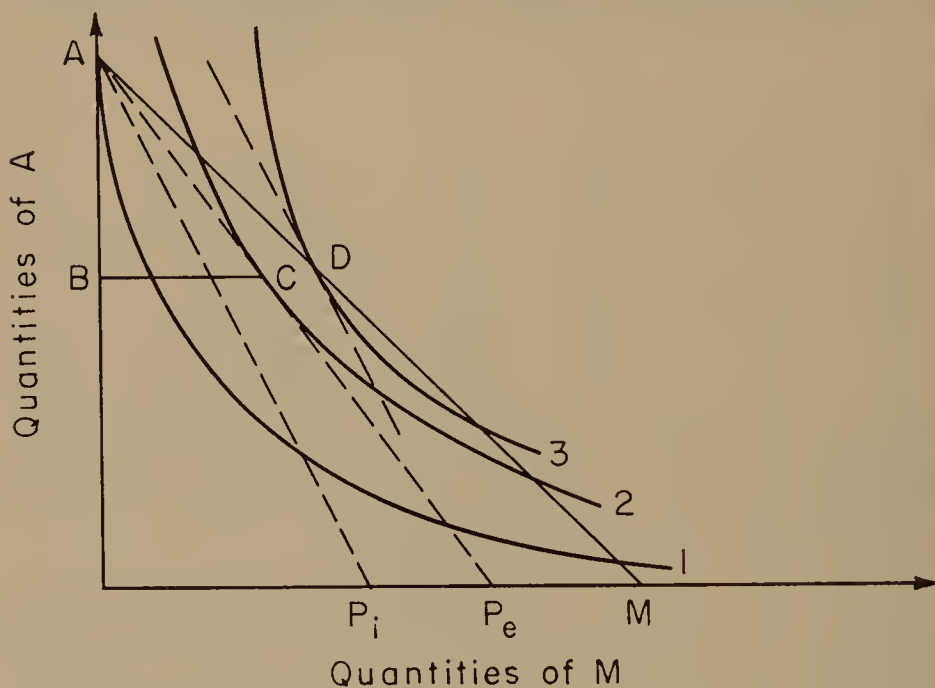


FIGURE I

FREE TRADE AND PROTECTIONISM IN THE ONE-FACTOR CASE

AM be the marginal transformation curve relative to apples and suits. Since there is only one factor, it is a straight line.⁵ But let wages in manufacturing be twice as high as in agriculture. Since the cost of labor is the only cost of production, the domestic price ratio $\frac{P_M}{P_A}$ will be twice as great, and the exchange ratio, the number of units of M exchanged for one of A , one-half as great, as the trans-

5. It will be concave upward at A , if there is a minimum size plant for the production of suits below which size production is inefficient. Similarly at M for the production of apples. As is usual in drawing transformation curves, I ignore this possibility both here and in Figures II and III.

formation ratio. Let AP_i (or the dashed line parallel to it through D) be this domestic exchange ratio. And let the external exchange ratio, AP_e , be between AP_i and AM in slope. The numbered curves concave upward are community indifference curves. With due recognition of the inaccuracy of assuming one set of such curves, because a change in the composition of output changes the distribution of income and hence affects the shape of the curves, I use a single set for convenience in exposition. Since it is reasonable to assume continuity in shifts in the curves, no plausible assumption about shifts in them associated with changes in the composition of output will alter the conclusions reached here.

If free international trade is permitted, no one will buy suits for apples at home, because imported suits are cheaper in terms of apples. The economy will specialize in the production of apples; i.e., production will be at the point A . Since consumers are willing to exchange apples for suits at a ratio equal to the slope of indifference curve 1, they will gladly exchange domestically produced apples for imported suits at the exchange ratio AP_e . AB of apples will be exchanged for BC of suits, bringing the economy to indifference curve 2, the highest curve that can be reached by this trading.

Suppose now that international trade is prevented, by prohibitive tariff or other device. Indifference curve 1 indicates that some consumers are willing to trade apples for suits on even worse terms than the domestic exchange ratio AP_i . They will therefore offer to do so at that ratio, and producers will turn to the production of suits. As they do so, they will however move, not along the exchange ratio line AP_i , but along the transformation curve AM . They will move along that curve as far as the demand for suits at a price equal to or above the exchange ratio AP_i persists, i.e., until that price ratio is tangent to a community indifference curve, as at point D . This is the equilibrium point, for consumers would exchange more apples for suits only at an exchange ratio more favorable to suits; and will exchange suits for apples only at a rate more favorable to apples. At point D , on indifference curve 3, welfare is greater than could be attained by international trade.

The two-factor case

Before drawing implications from this conclusion, let us turn to the two-factor case. In Figure II, ARM is the "efficiency locus," i.e., the transformation curve between A and M assuming perfect competition in all markets and hence identical price in both industries for the units of any input. If now we assume that wages are higher

in manufacturing than in agriculture, then "too little" labor will be used in manufacturing because of its high price, and, capital being relatively cheap, "too much" capital will be used. The opposite is true in agriculture. Production will be less efficient than before.

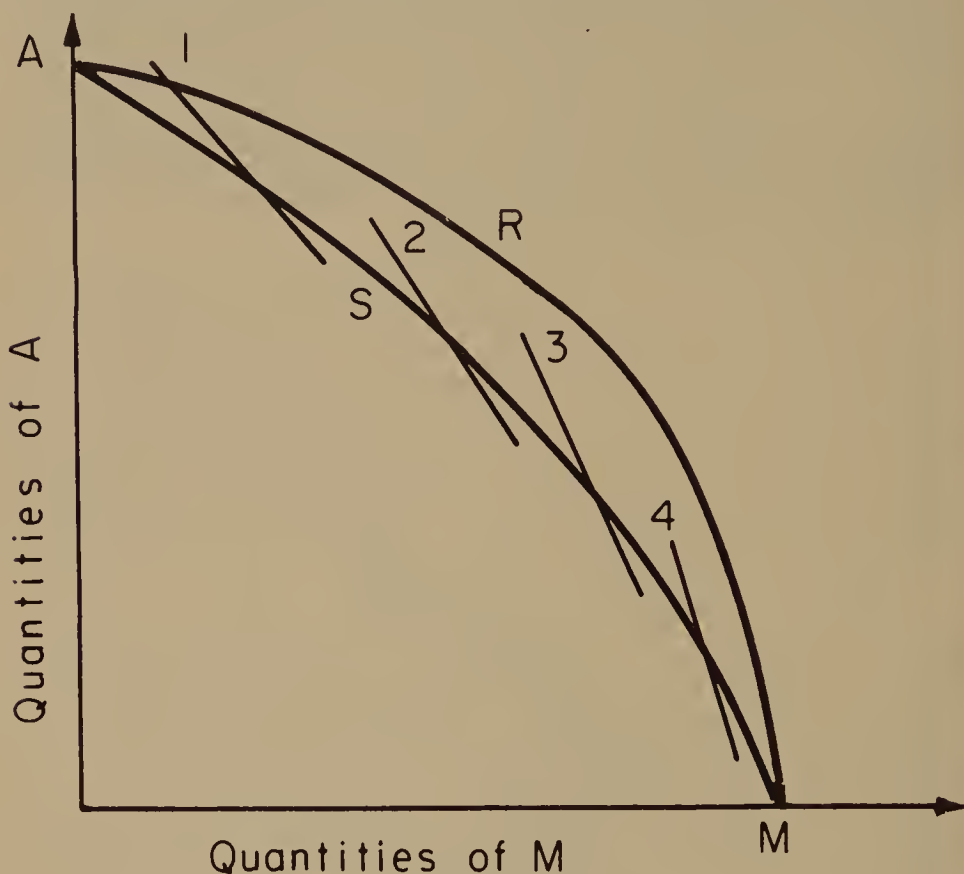


FIGURE II

TRANSFORMATION CURVE AND EXCHANGE RATIOS WITH A WAGE DIFFERENTIAL

The transformation curve which embodies this wage constraint will lie inside the curve ARM .⁶ In Figure II it is represented by

6. This is intuitively obvious. It can be proved rigorously by use of the Edgeworth-Bowley box diagram from which the locus of an economy's transformation curve is derived. (See Stolper and Samuelson, *op. cit.*) In this diagram, each production possibility curve of A is tangent to a production possibility curve of M . The efficiency locus is the locus of all of these points of tangency. With the wage constraint, however, the transformation curve is the locus of points of intersection of pairs of production possibility curves at each of which their slopes differ by a definite amount determined by the wage differential. At any given point on this locus, for a given output of either product, less is produced of the other than on the efficiency locus; i.e., any given production-possibility curve of either product intersects a lower curve of the other than the curve to which it is tangent.

ASM. That the two curves coincide at the extremities results from the assumption of full employment implicit in the Edgeworth-Bowley box from which they are derived. This is appropriate in the present model. It should be noted that there is no inflexibility of wages, but only a differential. If excess labor threatens to appear (in agriculture), its competition will cause wages to fall in agriculture and more labor to be used there. At the same time, the increased wage differential between agriculture and manufacturing will increase the flow of labor from agricultural areas to manufacturing, bringing wages down there also and causing labor absorption there. Thus wages in both industries are flexible relative to capital costs, but as they move the interindustry wage differential is maintained by the dynamics of the system.⁷

Lines 1, 2, 3, and 4 are exchange ratios between A and M determined by relative marginal costs of production. Under conditions of perfect competition, the exchange ratios are tangent to the transformation curve. That is, the exchange ratio is equal to the marginal rate of substitution between A and M (marginal transformation ratio). But if wages of equivalent labor are higher in industry than in agriculture, the reciprocal of the marginal money cost ratio is no longer equal to the marginal transformation ratio, and the exchange ratio line will cut the transformation curve as shown, the exchange ratio $\frac{1}{P_A/P_M}$ having a steeper negative slope at any point than $\frac{dA}{dM}$, the slope of the curve.⁸

7. If unemployment should appear in manufacturing, wages will fall there, causing labor absorption; and at the same time the shrinking of the interindustry wage differential will lessen the flow of labor from agriculture, thus causing a fall in wages and absorption of labor there (and maintaining the wage differential).

8. Since the average cost per unit of inputs in producing M is higher than of equivalent inputs in A , intuitively it seems likely that the price ratio $\frac{P_A}{P_M}$ would be

lower than the marginal transformation ratio $\frac{dA}{dM}$, and hence that the number of

units of A which will exchange per unit of M is greater than $\frac{dA}{dM}$. The proof is as follows:

Let $\frac{X_A}{X_M}$ be the exchange ratio between A and M , that is, the number of units of A

that will exchange for one unit of M .
$$\frac{X_A}{X_M} = \frac{1}{P_A/P_M}.$$

Let W = the wage rate, r = the rate of return on capital, and subscripts indicate the industry or output referred to.

To prove that $\frac{X_A}{X_M} > \frac{dA}{dM}$

We start with the conditions:

$$A = A(L, K) \quad M = M(L, K)$$

$$r_M = r_A = r \quad W_M = \alpha W_A, \text{ where } \alpha > 1.$$

$$dA = \frac{\partial A}{\partial L_A} dL_A + \frac{\partial A}{\partial K_A} dK_A \quad (1)$$

$$dM = \frac{\partial M}{\partial L_M} dL_M + \frac{\partial M}{\partial K_M} dK_M \quad (2)$$

$$P_A = \frac{\partial L_A}{\partial A} W_A = \frac{\partial K_A}{\partial A} r \quad (3)$$

$$P_M = \frac{\partial L_M}{\partial M} \alpha W_A = \frac{\partial K_M}{\partial M} r \quad (4)$$

$$\frac{dA}{dM} = \frac{\frac{\partial A}{\partial L_A} dL_A + \frac{\partial A}{\partial K_A} dK_A}{\frac{\partial M}{\partial L_M} dL_M + \frac{\partial M}{\partial K_M} dK_M} \quad (5)$$

$$\frac{P_A}{P_M} = \frac{\frac{\partial L_A}{\partial A}}{\alpha \frac{\partial L_M}{\partial M}} = \frac{\frac{\partial K_A}{\partial A}}{\frac{\partial K_M}{\partial M}} \quad (6a)$$

$$\frac{P_M}{P_A} = \frac{\frac{\partial A}{\partial L_A}}{\frac{1}{\alpha} \cdot \frac{\partial M}{\partial L_M}} = \frac{\frac{\partial A}{\partial K_A}}{\frac{\partial M}{\partial K_M}} \quad (6b)$$

Or, multiplying numerator and denominator by the same amount

$$\frac{P_M}{P_A} = \frac{\frac{\partial A}{\partial L_A} dL_A}{\frac{1}{\alpha} \cdot \frac{\partial M}{\partial L_M} dL_M} = \frac{\frac{\partial A}{\partial K_A} dK_A}{\frac{\partial M}{\partial K_M} dK_M} = \frac{\frac{\partial A}{\partial L_A} dL_A + \frac{\partial A}{\partial K_A} dK_A}{\frac{1}{\alpha} \cdot \frac{\partial M}{\partial L_M} dL_M + \frac{\partial M}{\partial K_M} dK_M} \quad (6c)$$

Since $\frac{1}{\alpha} < 1$,

$$\frac{P_M}{P_A} > \frac{dA}{dM}$$

$$\therefore \frac{P_A}{P_M} < \frac{dA}{dM}, \text{ and } \frac{X_A}{X_M} > \frac{dA}{dM} \quad Q.E.D.$$

That is, the number of units of A which will exchange for one unit of M is greater than the number of units of A whose production must be sacrificed to obtain production of one unit of M . Geometrically, the exchange ratio line 1 of Figure III is more steeply sloped (negatively) than the transformation curve $AP'M$, at any point on the curve.

Now turn to Figure III.⁹ As before, the curves concave outward are community indifference curves, and those concave to the origin are the two transformation curves. Line 3 is the external exchange ratio. Lines 1 and 2 are domestic exchange ratios, line 1 being that one that is identical with the external ratio. The external exchange ratio is assumed given, i.e., not affected by shifts in production and trade within the country being analyzed. (Relaxing this assumption, i.e., assuming a curved external terms of trade line, would not affect

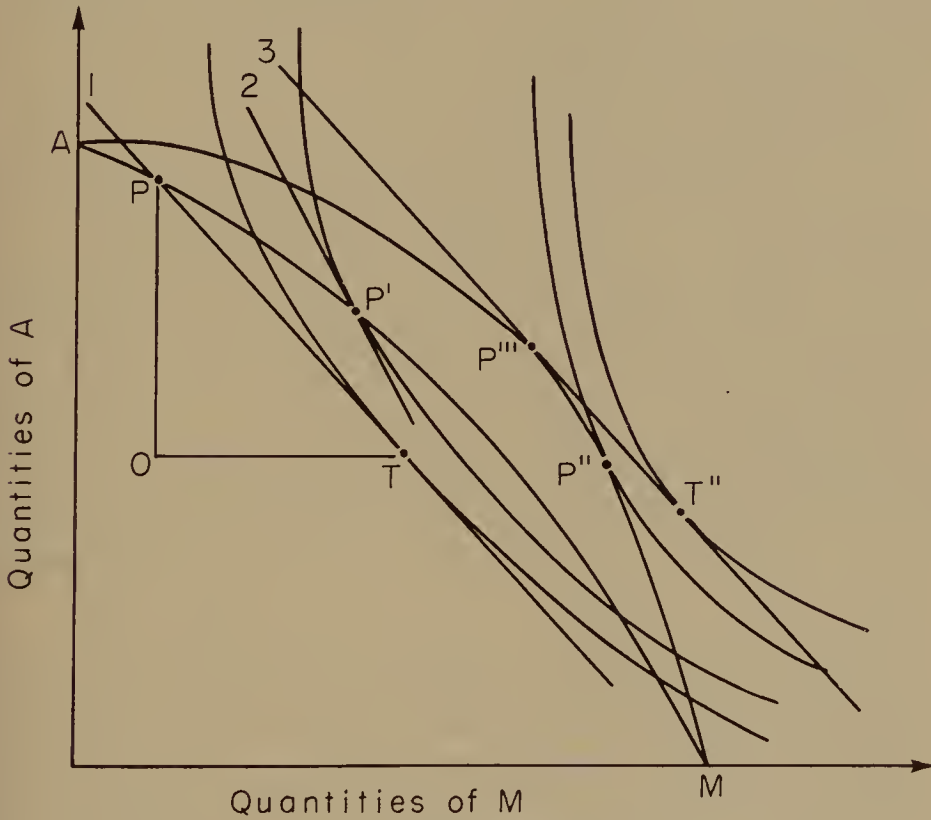


FIGURE III

FREE TRADE, PROTECTIONISM, AND SUBSIDY IN THE TWO-FACTOR CASE

the argument.) If there is free trade, the external exchange ratio will, of course, prevail in the country.

Production in the country will therefore settle at P , the point on $APP'M$ at which the external terms of trade line coincides with the internal exchange ratio (hence not at the point of tangency between the external exchange ratio and $APP'M$). For at any point

9. Though in his *Economic Journal* article, *loc. cit.*, Haberler has no curve such as $AP'M$, the geometry of Figure III is directly derived from the Haberler geometry.

to the left of P , the exchange ratio is flatter, i.e., a unit of M can be produced more cheaply relative to A , than the external terms of trade. Hence to maximize profits producers will increase production of M and decrease that of A . At any point to the right of P , the opposite will occur.

But though production will settle at P , consumption will not. For by exchange with the rest of the world at the prevailing terms of trade, shown by the slope of line 1, the economy can move to a higher indifference curve. Hence PC of A will be exchanged for CT of M , and the economy will reach T , the optimum point attainable under free trade in the given conditions.

Now consider the situation if a tariff or other protectionist measure prevents foreign trade. As the slope of the community indifference curve at P (not drawn) indicates, demand in the economy will bid up the relative price of M , and production will move along $APP'M$ to the point where the community indifference curve and the exchange ratio are tangent, namely P' .¹ But this is a point of higher real income than P . Protection of manufacturing from foreign trade will increase real income.

This, however, is not the end of the story. The economy is still inside the efficiency locus $AP'''P''M$. A subsidy per unit of M could move it farther along $APP'M$ toward M , but no measure which leaves the $\frac{W_M}{W_A}$ ratio unchanged can move it above or outside that curve.

It can, however, be moved to a point on $AP'''P''M$ by a different sort of measure. Assume that a subsidy per unit of labor in industry, equal to the difference in unit labor cost between agriculture and industry, is introduced. If so, the exchange ratio, $\frac{1}{P_M/P_A}$, will be

equal to $\frac{dM}{dA}$ and to $\frac{\partial M/\partial L_M}{\partial A/\partial L_A}$ and $\frac{\partial M}{\partial K_M} / \frac{\partial A}{\partial K_A}$. The transformation

curve will therefore move out to $AP'''P''M$, and output will be at some point on that curve. If the prohibitive tariff still prevails, output will settle at point P'' , the point of tangency between $AP'''P''M$ and a community indifference curve² and therefore the optimum point attainable without trade. But if the tariff is removed,

1. The wage differential will be increased temporarily by the shift of production from P to P' , which requires industry to recruit added workers. The exchange ratio may therefore be steeper than line 2. If so, the point of production will temporarily be to the left of P' .

2. Subject to a temporary effect parallel to that stated in the preceeding footnote.

the exchange ratio between A and M will move to identity with the external terms of trade (line 3, parallel to line 1), production will settle at point P''' , and by trade the system will move to the still more advantageous point T''' .

(Similarly in Figure I a subsidy per unit of labor would move production to point M , and by trade the system would move to an indifference curve higher than curve 3.)

Extensions and qualifications

Perhaps it is worthwhile to note that the effects sketched will occur whether industry is more or less labor-intensive than agriculture, in any of the several possible meanings of the term *labor-intensive*. They occur not because of differences in factor proportions, but purely because of the factor cost differential.

And perhaps it should also be noted, in conclusion, that since an interindustry differential in the cost of any factor has the effect sketched here, and since interest rates are typically higher in agriculture than in industry, there is in life a counterweight to the effect sketched here.³ However, interest costs are typically much smaller than labor costs, and specifically the differential in interest costs is typically a far smaller share of total costs than the differential in wage costs. Hence the empirical implications of the argument sketched here are not greatly reduced by the existence of the counterweighting interest differential.

The meaning of the conclusion reached may be summarized as follows. Let us express all incomes and prices in terms of the value of one unit of agricultural product, in order to avoid price level problems. In these terms, protectionism raises real income, relative to free trade, if the increase under protection in the aggregate cost of the industrial product to its buyers is less than the increase in income to the factors which shift from agriculture to industry. (This is equivalent to stating that real income will be increased by protectionism if, assuming factor costs identical in agriculture and industry, the economy could produce the industrial product at a lower cost,

3. A counterweight in so far as the relative price of agricultural to industrial products is concerned. By moving the price ratio toward the value it would have under perfect competition, the interest differential tends to move the system toward the "proper" point on the production possibility curve. In another sense, the interest differential accentuates the effect of the wage differential. By making the factor proportion in each industry even farther from the optimum than it would be on account of the labor differential alone, the interest differential reduces even further the efficiency of production, i.e., causes the system to move to a production possibility curve even inferior to that one resulting from the wage differential alone.

expressed in units of agricultural product, than the import price.)

To conclude this analytical section, let me comment briefly on the case of a many-product system and on the question of world versus national welfare. For convenience, I shall consider the many-product system to be composed of three sectors, agriculture, industry, and services. The simplification will not distort the argument.

As per capita income rises, the service sector will grow relative to industry as well. However, the nontransportability of services from abroad provides domestic service industries with a "natural" equivalent of protection. In terms of Figure III, they are always at P' rather than at P . Hence, protection to manufacturing industry creates no distortion as between manufacturing and service industries, and increases welfare, just as in the two-product model.

The effect of subsidies to manufacturing alone is more ambiguous. For if wages in service industries for equivalent labor are as high as in manufacturing, then a subsidy to manufacturing per unit of labor equal to the agriculture-manufacturing wage disparity, will distort the allocation of inputs between manufacturing and services even while it removes the distortion in the allocation between agriculture and manufacturing.

Empirically, it may be questioned whether the agriculture-service wage disparity is as great as that between agriculture and manufacturing, since in general the difference between rural work and that in the expanding service industries is probably much less than that between agriculture and manufacturing, hence mobility is greater. It may also be suggested that in the early stages of development the relative rate of growth in the tertiary service industries is much less than later, hence the two-product case is more clearly applicable. These, however, are empirical questions, and can be settled, if at all, only by empirical research. The only safe simple analytical observation is that the *optimum optimorum* can be reached only by subsidizing all wage differentials resulting from growth.

Protectionism that increased national welfare might in a number of obvious cases diminish world welfare, for (assuming a wage differential in all countries) protection that increased real income in the country previously importing industrial products would in a typical case reduce real income in the country previously exporting industrial products. But free trade without an offset to the factor price disequilibrium is not the optimum situation from the world viewpoint, for it would leave the world on the inefficient transformation curve $APP'M$. Only a labor cost subsidy in every country, combined with free trade, would bring the world to the *optimum optimorum*.

IV. EMPIRICAL COROLLARY: PROTECTIONISM AND GROWTH

The broad historical record suggests that protectionism may have accelerated economic development. A number of countries that have entered upon economic development since the original industrial revolution in England have done so behind a protectionist wall. This is true, for example, of the United States, Japan, the Soviet Union,⁴ and the three Latin American countries — Brazil, Colombia, and Mexico — whose per capita income is now rising rapidly. In every one of these cases, the rise in per capita income was associated with a sharp increase in the share of total income originating in manufacturing industry. Income originating per person engaged in industry was clearly higher than in the economy as a whole, and in a simple statistical sense industry contributed greatly to the rise in per capita income. If the effect had been only an infant industry one, the rise in income would not have occurred until the industries became viable without tariff protection. In fact, however, it has occurred while they were unable to exist without the tariff.

Now it is possible that the apparent contribution of industry was not real. The higher per capita income in industry was of course accompanied by an increase in the price of industrial products above the imported price. It is possible that if industrial protection had not existed, the entrepreneurs who started industrial ventures would have devoted their energies to other ventures, and would have drawn capital and labor so effectively into other ventures that real per capita income would have increased even faster than it did under protectionism.

It is possible, but it does not seem probable. For the rise in both aggregate and per capita income in protectionist countries possessing no notable resource advantages — Japan and the three Latin American countries, for example — has been rapid, relative to rates of economic growth elsewhere. On the basis of the conventional analytical model, economists have usually assumed either that income rose in spite of protection, or that protection increased income only through an infant industry effect or because of external economies.⁵

4. In the sense that the Soviet Union was isolated from foreign commerce by autarkical state controls. The United States had high tariffs from 1818 to 1833, and from the Civil War on. France and Germany had a mixed tariff history during their periods of industrialization. France had high tariffs from 1808 to 1860, and Germany from 1844 to 1860 and after 1879. The period of low tariffs after 1860 was a period of rapid economic growth in both countries; but development must have begun and taken root before 1860, and Germany's growth after 1879 was also very rapid.

5. Note, however, the implication of Viner's argument. If he accepted the

The model presented here suggests however that the apparent historical phenomenon may be accepted at face value and without need to resort to these explanations.

CENTER FOR INTERNATIONAL STUDIES
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

reality and necessity of a wage differential, his logic, if pursued, should lead to the conclusions presented here. And see W. Arthur Lewis, *op. cit.*, pp. 185-86. He advances an argument for protectionism, based on the assumption of the existence of "surplus" labor in agriculture whose real marginal cost for use in manufacturing is zero. This is a specific case of the general argument presented here.

A STUDY OF DECISION-MAKING WITHIN THE FIRM*

By EDWIN MANSFIELD AND HAROLD H. WEIN

I. Freight yards, 516 — II. Output and employment decisions, 517. — III. Switching output, 518. — IV. Delivery output, 523. — V. Aversion functions, 526. — VI. Costs and the complete model, 528. — VII. Tests and prediction, 529. — VIII. Summary and conclusions, 535.

For various reasons, economists have expressed dissatisfaction with the existing theory of the firm.¹ One such reason is the failure of the theory to take greater account of intrafirm variables and organizational relationships. The "economic firm," composed principally of a dominant entrepreneur and some passive factors of production, is of course a convenient simplification. Some economists, feeling that this simplification overlooks some important factors, have suggested models that focus more attention on the internal organization and workings of the firm.²

When the "black box" within the firm is opened, many interesting problems come to light. One such problem relates to the motives and decisions of managers at intermediate and lower levels of the organization. There can be no doubt that important decisions influencing the behavior of the firm are often made at such levels. And it would seem that some study can profitably be devoted to this area.

In this paper an attempt is made to explain some decisions of a manager at an intermediate level in a railroad and to link these decisions to the short-run fluctuations in output and cost of individual freight yards. The model that results seems useful in explaining the

* This report is based on research supported by a grant from the Westinghouse Air Brake Corporation to Carnegie Institute of Technology. It is part of a larger project concerning the railroad industry. An earlier version of the paper was read before the December 1956 meetings of the Econometric Society. We should like to acknowledge the valuable comments of our colleagues, particularly W. W. Cooper, J. Dreze, C. Haberstroh, and Franco Modigliani, and the assistance of C. Link, R. Nadel, and E. Saunders of the co-operating railroad.

1. See for example W. W. Cooper, "A Proposal for Extending the Theory of the Firm," this *Journal*, LXV (1951); R. Cyert and J. March, "Organizational Factors in the Theory of Oligopoly," this *Journal*, LXX (1956); K. Boulding, "Implications for General Economics of More Realistic Theories of the Firm," *American Economic Review*, XLII (1952); A. Papandreou, "Some Basic Problems in the Theory of the Firm," *A Survey of Contemporary Economics*, Vol. II.

2. For example, Cooper, *op. cit.*, Cyert and March, *op. cit.*, and Papandreou, *op. cit.* Of course, this does not imply that the traditional theory is not satisfactory for many purposes, e.g., as a basis for some types of aggregative analysis.

observed fluctuations in output and cost. This model is tailored in many ways for the industry and situation we study, but we believe that it affords some insight into the general problem. The following points might also be noted: (1) The model is not based on the usual assumption of profit maximization or cost minimization. (2) The model takes some account of the uncertainty under which decisions are made. (3) The paper contains the results of some empirical tests to which the model has been subjected.

The plan of the paper is as follows. Section I contains some introductory material concerning freight yards. This material is included for those readers who are relatively unfamiliar with the railroad industry. Sections II–VI contain a description of the model. Section VII is concerned with the testing of the model. Section VIII contains a summary and some concluding remarks.

I. FREIGHT YARDS

The manager with whom we are concerned is the general yardmaster of a freight yard. The general yardmaster operates under a broad grant of authority, and he makes basic decisions concerning output and employment at his freight yard. To understand these decisions fully, it is necessary to know something about the nature and function of freight yards.

Physical Characteristics. All yards contain sets of tracks. In large yards, they are generally of three types: receiving tracks where incoming freight cars are stored, classification tracks where cars are switched, and outbound tracks where cars that have been switched are stored until a locomotive hauls them away as a train.

Functions. Freight yards switch cars. That is, they sort incoming cars by putting them on the appropriate classification tracks, and in this way they break up incoming trains to form new trains. Most yards also deliver and pick up cars. Engines are assigned to deliver cars to industrial sidings and other yards and to pick them up there. Finally, many yards bill and inspect freight cars and perform such ancillary services as maintenance, repair, and storage.

Importance. A large railroad may have as many as 200 freight yards. In the firm we studied, the operation of these yards accounts for about one-third of total operating costs.³

Three terms relating to the switching process should be defined. (1) Unswitched cars that are present in the yard are called *backlog*,

3. For a more complete description of yard operations, see M. Beckmann, C. McGuire, and C. Winston, *Studies in the Economics of Transportation* (New Haven: Yale University, 1956).

and they correspond in a loose way to unfilled orders in manufacturing. (2) A *cut* is a group of cars that rolls as a unit onto the same classification track. (3) The *cut size* is the number of cars in such a group. All other things equal, a string of cars with a small average cut size will be processed more slowly than one with a larger average cut size.

Yard output should also be defined. Since the two most important services performed at the yard are switching and delivery, we shall use the number of cars switched and the number of cars delivered as an output measure. This seems to be a reasonable first approximation, but its crudeness should be recognized. In particular, switching and delivery are not homogeneous; for example, two groups of cars may be delivered but one group may be hauled a greater distance than the other.⁴

II. OUTPUT AND EMPLOYMENT DECISIONS

Let us return now to the yardmaster. Note first that his decisions concerning output and employment are made within a somewhat different framework than that which prevails in manufacturing. (1) Because of the nature of the output, he cannot produce for inventory. (2) He cannot reject "orders." Under normal circumstances, he must service all cars that enter his yard, but he can generally decide when and how they are serviced. (3) The planning period is very short, and his decisions concerning output and employment refer generally to only the next twenty-four hours.

Next, note the decisions the yardmaster makes. Each day he plans the number of cars to be switched and the number of switch-engine crews to be employed on the next day. He also plans the number of cars to be delivered and the number of delivery crews for the next day. Because they have so strong an influence on the yard's performance and cost, these decisions concerning output and employment are very important. Our purpose here is to explain these decisions, in particular the output decisions, and to link them with observed output levels and costs.

Two further points should be noted. (1) The decisions treated here are at face value only tentative. It is often possible for the yard-

4. An alternative measure, used in E. Mansfield and H. Wein, "A Regression Control Chart for Costs," *Applied Statistics* (1958), is the number of cuts switched and the number of cars delivered. The railroads generally use the number of cars handled as an output measure. See E. Mansfield and H. Wein, "Notes on Railroad Productivity and Efficiency Measures," *Land Economics* (1958). Both cuts switched and cars switched could be included in an output measure but for our purposes it seemed reasonable to include only one because they are so highly correlated.

master to change them during the next day. However, it appears that these plans are seldom modified appreciably and that they are actually quite firm.⁵ (2) When we say that he plans the number of cars to be switched or delivered on the next day, we mean that he plans the number of cars that *can* under normal circumstances be switched or delivered then. He plans the switching or delivery capacity; whether or not this capacity is fully used depends on how many cars arrive and how they arrive. This point is discussed further below.⁶

In the next section, a simple model is constructed to explain the number of cars the yardmaster plans to switch and these plans are linked with the actual switching output. In Section IV, we treat the number of cars he plans to deliver, and these plans are linked with the actual delivery output. By our definition of yard output, the combined results of these two sections constitute a model explaining actual yard output. In Section VI, yard costs are taken up, and the results of previous sections are combined to produce our complete model to explain actual yard output and costs.

III. SWITCHING OUTPUT

The yardmaster plans each day the switching output and the number of switch-engine crews for the next day. We suppose that these plans are formulated on the basis of his forecast of the next day's

5. The output and employment plans for the next day can be changed during the next day. With regard to employment, it is commonly possible for extra crews to be hired or laid off two hours or more before a shift begins. However, the yardmasters with whom we spoke seemed to believe that changes seldom occur and that when they do, they are small. The little evidence that is available seems to indicate this too:

Data for thirteen days were collected at one yard concerning the planned and actual number of switch-engine crews. On seven days, the planned and actual number of crews were the same. On two days there was a discrepancy of 4 per cent; on four days there was a discrepancy of 8 per cent. This was a period when the yardmaster's traffic forecasts were relatively poor, and it is reasonable to believe that the discrepancies would ordinarily be smaller.

Data concerning the planned and actual switching outputs were collected for twelve days at one yard and for thirteen days at another. On ten days the discrepancy between planned and actual output was 5 per cent or less; on only five days the discrepancy was 10 per cent or more. Moreover, it is likely that the discrepancies might ordinarily be smaller.

Data for thirteen days at one yard indicate that on only three days did the planned delivery output differ from the actual output by more than eighty cars.

6. The distribution of train arrivals over time is important here. We assume that he thinks in terms of some average distribution of train arrivals while planning. Note that "capacity," as used here, has quite a different meaning than it does later in the paper when we speak of the capacity of a yard.

switching traffic.⁷ (The way in which his plans depend on this forecast is discussed below.) The switching traffic is the number of cars that are available for switching, and we suppose that

$$W_i = t_i + b_i - f_i, \quad (1)$$

where W_i is his forecast of the switching traffic for the i^{th} day, t_i is his forecast of the number of cars that will arrive on the i^{th} day to be switched, b_i is the backlog he expects at the end of the $i-1^{\text{th}}$ day, and f_i is the *forced backlog* he expects on the i^{th} day. The forced backlog is the number of cars that arrive so near the end of the period that they cannot possibly be switched before the period ends.

We assume next that both the backlog and the idle crew time that the yardmaster foresees for the next day are determined by the discrepancy he foresees between the actual switching traffic and the output he plans. We assume that he believes that (1) the actual and planned number of crews will be equal, (2) the actual and planned outputs will be equal unless the switching traffic falls short of the planned output, (3) the backlog will equal the forced backlog if the switching traffic falls short of the planned output, (4) there will be little or no idle crew time if the switching traffic exceeds the planned output, and (5) the crew productivity is a known constant if they are fully employed.⁸ Then the amount of backlog he foresees for the next day equals a constant plus the excess of the switching traffic over the planned output.⁹ And the amount of idle crew time he foresees

7. Most yardmasters seem to make fairly explicit forecasts for each succeeding day. Whether or not they formulate a forecast in precisely the way we suppose is unknown.

8. Note that these assumptions relate to what he tentatively believes or figures on when planning, and not to what will actually occur. We may reasonably expect him to believe that the planned and actual employment (and output) will be equal where possible. Moreover, given the distribution of train arrivals that he expects, it does not seem unreasonable for him to make assumptions (3), (4), and (5). By crew productivity being constant, we mean that it does not vary with the number of crews. Note, too, that these assumptions imply that, while planning, he does not take account of whatever possibility he has of changing his plans.

9. The actual switching traffic, S_i , equals $B_i + T_i - F_i$, where B_i is the *actual* backlog at the end of the $i-1^{\text{th}}$ day, F_i is the *actual* forced backlog on the i^{th} day, etc. Since $B_{i+1} = B_i + T_i - Q_i$, where Q_i is the actual output on the i^{th} day, it follows that:

$$\begin{aligned} B_{i+1} &= S_i - Q_i + F_i \\ &= S_i - Q_i + F, \end{aligned}$$

if we assume F_i to be constant. Since the planned and actual outputs are supposed to be equal, the backlog he foresees for the next day equals a constant plus the excess of the switching traffic over the planned output. Assumption (3) above covers the case where the planned output exceeds the switching traffic.

is proportional to the excess of the planned output over the switching traffic.¹

Next, we note that the yardmaster will incur penalties if (1) a large backlog exists at the end of the next day and serious congestion results, or if (2) the idle crew time and hence the cost per car switched is high on the next day. We refer to congestion. As the number of cars in the backlog increases, the yard becomes more congested. Up to a point, increases in backlog result in little difficulty; but beyond that point, congestion becomes serious and there is a reduction in crew productivity and an increase in car delay. We speak too of penalties. Penalties should not be confused with the costs the railroad incurs. They are meted out by the yardmaster's superiors, and they often take the form of reprimands or transfer to a lesser position. They may result from decisions that are costly to the railroad, and they need not be strictly related to costliness.²

The yardmaster is supposed to attach certain "disutilities" to the penalties he incurs. Assuming that the penalties depend uniquely on the amounts of backlog and idle crew time,³ he in effect attaches these "disutilities" to the latter. When he plans the next day's output, each discrepancy he foresees between the switching traffic and the planned output means a certain amount of backlog and idle crew time on the next day. Hence, when he plans, he attaches a certain "disutility" to each possible discrepancy. In choices involving uncertainty, we suppose that these "disutilities" can be described by a cardinal aversion function like that in Figure I.⁴

The function in Figure I is, of course, only a convenient approxi-

1. We assume that the planned number of crews equals $\frac{X_i}{P}$ where X_i is the planned output and P is the crew productivity. This ignores discontinuities in the possible number of crews. The number of crews necessary to switch the actual output (i.e., the switching traffic) is S_i/P , where S_i is the switching traffic. The prospective idle time is therefore $\frac{X_i - S_i}{P}$. Assumption (4) covers the case where $S_i > X_i$. Day-to-day changes in P could readily be taken into account but we ignore them here.

2. For example, a decision may be costly to the railroad, but if his superiors are unaware of the costs or if they believe that he is not responsible, the yardmaster may incur no penalty.

3. Of course, this is a simplifying assumption. Actually, for each amount of backlog and idle time there is some uncertainty concerning the penalty incurred. We assume for simplicity that his belief concerning the penalty resulting from each prospective amount of backlog and idle crew time is correct.

4. The conditions under which a cardinal preference function, and hence a cardinal aversion function, exists are stated in J. Marschak, "Rational Behavior, Uncertain Prospects, and Measurable Utility," *Econometrica*, XVIII (1950).

mation. However, its general shape seems quite reasonable. A discussion of the shape of the function is postponed to Section V.

We assume next that the actual switching traffic on the next day is a random variable that the yardmaster believes to be normally distributed. Then there corresponds to each output that he might plan (1) a probability distribution of discrepancies between the switching traffic and the planned output and (2) an expected "disutility." We assume that he plans the output for the next day that minimizes the expected "disutility."

The expected "disutility" (divided by $A(-L_2)$, a constant) is:

$$E(D) = \int_{-\infty}^{x_i - L_2 - E_i} \frac{p(t)}{\sigma} dt + \alpha \int_{\frac{x_i + L_1 - E_i}{\sigma}}^{\infty} p(t) dt, \quad (2)$$

where E_i and σ are the expected value and standard deviation of the switching traffic on the i^{th} day, $p(t)$ is the unit normal probability

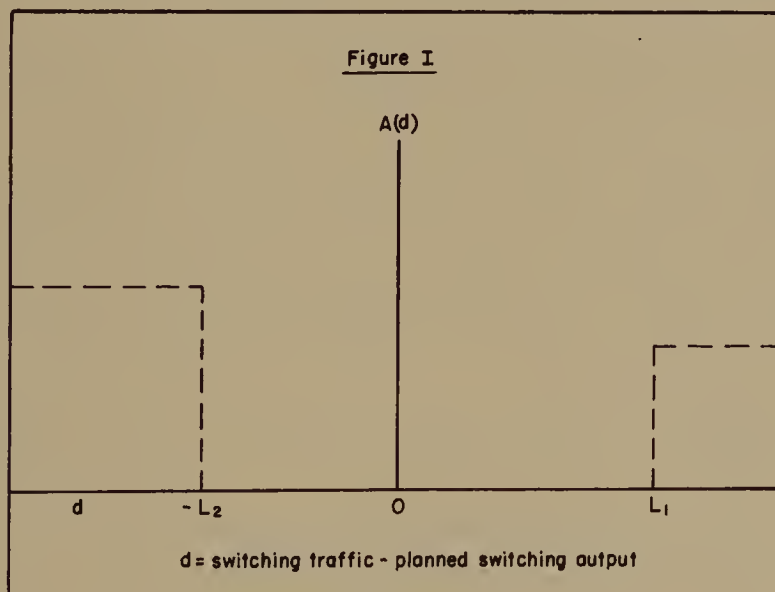


FIGURE I

density function, x_i is the output planned for the i^{th} day, and α is the ratio of $A(L_1)$ to $A(-L_2)$. Minimizing (2) with respect to x_i , we find that

$$x_i^* = E_i + \frac{\sigma^2 \log \alpha}{L_1 + L_2} + \frac{L_2 - L_1}{2}, \quad (3)$$

where x_i^* is the output planned for the i^{th} day that minimizes (2).

Assuming that the yardmaster's forecast of the switching traffic equals E_i ,

$$x^*_i = b_i + t_i - f_i + c, \quad (4)$$

where c is the sum of the last two terms on the right in (3). Assuming f_i to be constant from day to day,⁵

$$x^*_i = b_i + t_i + c'. \quad (5)$$

Hence, the switching output planned for the i^{th} day equals the amount of backlog expected at the end of the $i - 1^{\text{th}}$ day plus the forecasted number of cars that will arrive to be switched on the i^{th} day plus a constant. To link this planned output with the actual output, we assume that the planned and actual outputs correspond quite closely. More precisely, we suppose that

$$s_i = x^* + u_i = b_i + t_i + c' + u_i, \quad (6)$$

where s_i is the actual number of cars switched on the i^{th} day and u_i is a random variable with zero expected value and small variance. The data that are available seem to be compatible with this assumption.⁶

Note that there is no contradiction between (1) this assumption of close correspondence between planned and actual outputs and (2) the recognition that on some days the planned output may exceed the switching traffic and hence that the planned and actual outputs cannot be equal. Because c appears generally to be quite large and negative, the yardmaster plans in such a way that the probability of the planned output exceeding the switching traffic by substantial amounts is small. The planned output may often exceed the switching traffic by small amounts; but the resulting discrepancies between planned and actual output can easily be represented (with discrepancies resulting from other factors) by the random variable, u_i .

Finally, three points should be noted concerning the model:

(1) The yardmaster is presumed to plan the output level without explicit regard for the production or cost functions. This may not be unreasonable in the case of freight yards. Variable unit costs are often believed to be fairly constant up to a capacity output level; and, at most yards that we studied, existing traffic seldom approaches capacity. The yardmaster believes he will seldom, if ever, incur penalties on these grounds.

5. Data for two yards seem to indicate that f_i does not vary much from day to day. The mean deviation of the yardmaster's forecasts of daily forced backlog was forty-one cars at one yard and fifty-five cars at another.

6. We tested the hypothesis that the expected value of u_i is zero. The results based on the data cited in footnote 5, Section II, did not indicate that the hypothesis should be rejected.

(2) In a case like this, there is also no reason to believe that the expected traffic subsequent to the i^{th} day should affect his decision on the $i - 1^{\text{th}}$ day concerning the i^{th} day. It seems legitimate to ignore traffic subsequent to the i^{th} day when discussing the decision.⁷

(3) According to the model, the yardmaster is concerned with each day's performance and not just with monthly or annual averages. This seems reasonable. His superiors seem to check his performance each day on the basis of daily records.

IV. DELIVERY OUTPUT

Each day the yardmaster plans the delivery output and the number of delivery crews for the next day.⁸ We assume that these plans are formulated on the basis of his forecast of the next day's *delivery traffic*, i.e., the number of cars that are both available for delivery and wanted by customers. The relationship between his plans and this forecast is discussed below. When he makes the forecast, he distinguishes between two types of cars — industrial cars and interchange cars. The former are delivered to local industries whereas the latter are delivered to other railroads. The distinction is made because some industrial cars are ready for delivery but not wanted by customers.⁹ In the case of interchange cars, the other railroads must accept all cars that are delivered to them.

7. If we accept the framework of F. Modigliani, "The Role of Anticipations in the Economy of the Firm and Their Use in Economic Analysis and Forecasting," Conference on Expectations, Uncertainty, and Business Behavior, Social Science Research Council, 1955, the traffic subsequent to the i^{th} day is conditionally irrelevant. That is, so long as this traffic is not large enough to push the yard to capacity, the optimum decision concerning the i^{th} day is unchanged no matter what the subsequent traffic may be. Note that this relies on the point made in (1) above (preceding paragraph) concerning the way production and cost functions are viewed.

When we refer to capacity here and in (1), we mean the capacity of the whole yard. In (1) above, we refer to the supposed (approximate) constancy of average variable costs. It may be objected that there are discontinuities in the cost curve where additional crews are added. But the effect of this is weakened somewhat by the use of overtime and it is not so important in large yards.

8. Delivery output includes (a) the cars delivered to sidings and other yards, and (b) the cars brought back from sidings and other yards. In this and subsequent sections, however, delivery output is meant to include only the former. At the beginning of the study, railroad officials assured us that these two components of delivery output were so highly correlated that a satisfactory explanation of fluctuations in one would allow one to predict fluctuations in the other quite easily. It seems doubtful that this omission should limit appreciably the usefulness of this section.

To our knowledge, the yardmaster's plans concerning delivery are not affected by his plans concerning switching. The two jobs are quite distinct at large yards. A delivery crew cannot generally be used for switching, and vice versa.

9. Customers may hold cars at the yard three days without charge.

The yardmaster, when he makes the forecast, is supposed to make the following assumptions:¹ (1) All industrial cars available for delivery on the i^{th} day but not wanted by customers will be wanted on the $i + 1^{\text{th}}$ day. (2) The proportion of available industrial cars not wanted by customers is constant from day to day. (3) Each day, industrial cars constitute the same proportion of the total number of cars that arrive to be delivered. (4) The number of cars that arrive to be delivered but arrive so late that they cannot be delivered (before the end of the day) is constant from day to day. On the basis of these assumptions, his forecast will be²

$$V_i = (1 - P\nu)A_i + P\nu A_{i-1}, \quad (7)$$

where V_i is his forecast for the i^{th} day, A_i is the total number of cars he expects to arrive on the i^{th} day to be delivered, ν is the proportion of all cars that are industrial cars, and P is the proportion of available industrial cars that are not wanted by customers.

We note next that the yardmaster will incur penalties if (1) the idle crew time for the delivery crews and hence the cost per car delivered is relatively high or (2) the delivery of many cars is delayed until the day following their arrival. His superiors want interchange cars delivered before midnight to avoid an extra per diem charge,³

1. Available data seem to be compatible with the second and third assumptions. At two yards, the mean deviations of the yardmasters' daily forecasts of the proportion in assumption (3) were .01 and .02. The mean deviation of the yardmaster's daily forecasts of the proportion in assumption (2) was .02 at one yard. No data are available concerning the first and fourth assumptions.

2. Let A_i^c = number of interchange cars expected to arrive on i^{th} day,
 A_i^n = number of industrial cars expected to arrive on i^{th} day,
 L_i^c = number of interchange cars expected to arrive on i^{th} day too late to be delivered on that day,
 L_i^n = number of industrial cars expected to arrive on i^{th} day too late to be delivered on that day,
 p = proportion of industrial cars not wanted by customers,
 ν = proportion of all cars that are industrial cars,
 V_i = total number of interchange and industrial cars that can be delivered on the i^{th} day and are wanted then,
 $A_i = A_i^c + A_i^n$.

Then:

$$\begin{aligned} V_i &= A_i^c + L_{i-1}^c - L_i^c + (1 - P)(A_i^n + L_{i-1}^n - L_i^n) + P(A_{i-1}^n + L_{i-2}^n - L_{i-1}^n) \\ &= A_i - P\nu(A_i - A_{i-1}), \text{ by assumptions (3) and (4) above,} \\ &= (1 - P\nu)A_i + P\nu A_{i-1}. \end{aligned}$$

3. The railroad must pay \$2.75 per day for each "foreign" car on its tracks, and if a car is on its tracks at midnight it must pay charges for the next day.

and they want industrial cars delivered before the end of the day to avoid complaints by customers.

The yardmaster, when planning for the next day, is supposed to believe that to each possible discrepancy between the actual delivery traffic and the planned output, there corresponds a certain number of delayed cars and a certain amount of idle crew time on the next day. In particular, he is supposed to believe that (1) the number of delayed cars will equal the excess of the delivery traffic over the planned output, (2) the amount of idle crew time will be proportional to the excess of the planned output over the delivery traffic, (3) there

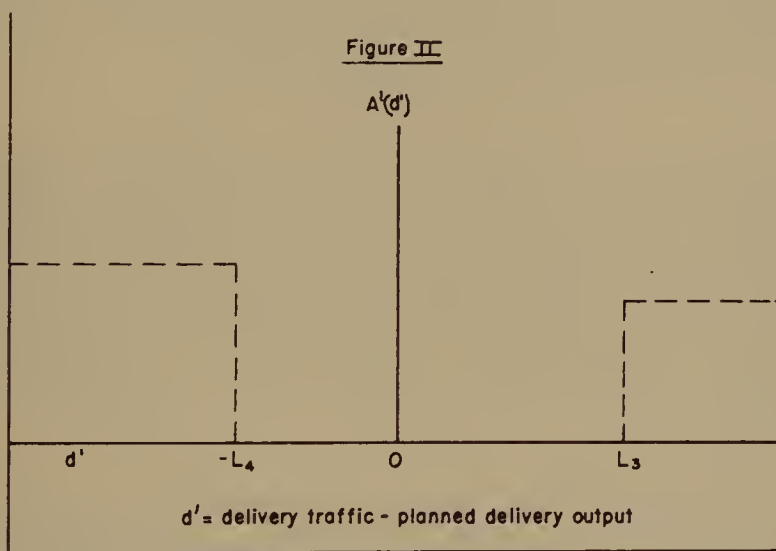


FIGURE II

will be no delayed cars if the planned output exceeds the delivery output, and (4) there will be little or no idle time if the delivery traffic exceeds the planned output.⁴

Next, we assume that there is a unique relationship between the amounts of car delay and idle crew time and the penalties he incurs. Then, assuming that he attaches certain "disutilities" to these penalties, he also attaches a certain "disutility" to each possible discrepancy between the delivery traffic and the planned output. In choices involving uncertainty, we suppose that these "disutilities" can be

4. It is sufficient that he believes (3) and (4) above and that (a) the actual employment on the next day will equal the planned employment, (b) where possible, the actual output on the next day will equal the planned output, (c) the crew productivity is known and constant, i.e., independent of the number of crews. The assumptions are similar to those in the previous section.

described by a cardinal aversion function like that in Figure II. (This function is discussed in the next section.)

The yardmaster is supposed to believe that the delivery traffic on the next day is a normally distributed random variable, and he is supposed to plan that output for the next day that minimizes the expected "disutility." Since the expression to be minimized is similar to (2),⁵ we know that

$$d^*_i = E'_i + \frac{\sigma'^2 \log \alpha'}{L_3 + L_4} + \frac{L_4 - L_3}{2}, \quad (8)$$

where d^*_i is the output planned for the i^{th} day that minimizes the expected "disutility," E'_i and σ'^2 are the expected value and variance of the delivery traffic on the i^{th} day, and α' is the ratio of $A'(L_3)$ to $A'(-L_4)$. Assuming that his forecast equals E'_i ,

$$d^*_i = (1 - P\nu)A_i + P\nu A_{i-1} + K, \quad (9)$$

where K is the sum of the last two terms on the right in equation (8).

Hence, the yardmaster plans to deliver on the next day a linear combination of the number of cars that he believes arrived to be delivered today and the number that he forecasts for tomorrow. To link this planned output with the actual output, we assume that there is close correspondence between them. In particular, we assume that

$$d_i = d^*_i + v_i = (1 - P\nu)A_i + P\nu A_{i-1} + K + v_i, \quad (10)$$

where d_i is the number of cars delivered on the i^{th} day and v_i is a random variable with zero expected value and small variance. The data that are available seem to be compatible with this assumption.⁶

V. AVERSION FUNCTIONS

In this section, we first describe some of the assumptions underlying the shape of $A(d)$ and $A'(d')$. Then we explain briefly why functions based on these assumptions seem to us to be reasonable approximations.

5. Merely substitute α^1 , σ^1 , E^1 , L_4 , and L_3 for α , σ , E , L_2 and L_1 in (2).

6. The hypothesis that the expected value of v_i equals zero was tested on the basis of the data mentioned in footnote 5, Sec. II. There was no indication that it should be rejected.

It should also be noted that many of the remarks concerning the switching model apply to the delivery model as well. For example, the last four points in the previous section are applicable.

Note, too, that we assume that α' (and α in the last section) is constant over time. This assumption need not be made. However, in the tests below, we use this simple assumption and the assumption that L_1 , L_2 , etc. are constant because no evidence exists to support another.

(1) Consider $A(d)$. For $d > 0$, the backlog foreseen for the next day equals d plus a constant, and hence it increases with d . We assume there exists an $L_1 (> 0)$ such that, for $d < L_1$, the yardmaster believes that the prospective backlog would be too small to produce serious congestion, his superiors would be unlikely to detect an appreciable deterioration in service, and he would incur no penalties. For $d \geq L_1$, we assume he believes that the prospective backlog would produce serious congestion, he would incur penalties, but these penalties would not vary much with the size of the backlog producing the congestion.

For $d < 0$, the prospective idle crew time is proportional to $-d$. We assume there exists an $L_2 (> 0)$ such that for $d > -L_2$, the yardmaster believes that the prospective idle crew time would not be great enough to exceed certain "tolerances" set for his performance by his superiors and he would incur no penalties. (In many cases, he believes that the prospective idle crew time would not even show up clearly on his superiors' records.) For $d \leq -L_2$, we assume he believes that the prospective idle crew time would be great enough to attract his superiors' attention and to exceed the "tolerance" interval, he would incur penalties for poor planning, but the latter would not vary greatly with the poorness of the planning.

(2) Consider $A'(d')$. For $d' > 0$, the prospective number of delayed cars equals d' . We assume there exists an $L_3 (> 0)$ such that, for $d' < L_3$, the yardmaster believes that the prospective number of delayed cars would not exceed certain "tolerances" set for his performance by his superiors and he would incur no penalties. (In many cases, he believes that the delayed cars may not even come to his superiors' attention since they have little direct information on this score.) For $d' \geq L_3$, we assume he believes that the number of delayed cars would result in enough complaints and other signals to prompt an investigation, the number would exceed the "tolerance" interval, he would incur penalties, but the latter would not vary much with the number of delayed cars.

For $d' < 0$, the assumptions are like those regarding $A(d)$ for $d < 0$.

The functions that are based on these assumptions are only approximations. To what extent do they seem reasonable? According to the yardmasters with whom we spoke, they seem quite reasonable. A more realistic function would have much the same general shape but it would be continuous. Apparently, there are certain points (L_1 to $-L_4$) up to which penalties are seldom imposed and beyond which they are usually imposed. In the case of $A(d)$ when

$d > 0$, this is because serious congestion is reached rather suddenly. In the other cases, it is due in part to deficiencies in the information channeled to the yardmaster's superiors and in larger part to certain "tolerances" they set up. The assumption that the penalties do not vary much beyond these points is reasonable for quite a wide range beyond them but not true for an indefinite range. Finally, the aversion functions depend on the yardmaster's subjective evaluation of the penalties as well as on the penalties themselves. From what we could gather, the approximations seem fairly reasonable when this factor is included.⁷

VI. COSTS AND THE COMPLETE MODEL

At this point, we have constructed a simple model to explain the yardmaster's output decisions and we have linked them with actual output. In this section, we add an equation to explain yard costs and we present the complete model to explain short-run fluctuations in output and cost.

On a given day, the money costs⁸ at a freight yard are a function of both the planned and actual output. Some factors were hired on the basis of the yardmaster's plans and they cannot easily be laid off or supplemented. The quantity of other factors that is used is almost entirely a function of actual output. We shall treat the costs as a function of actual output only. If the correspondence between planned and actual output is as close as we suppose, little harm should result from the omission of planned output.⁹ The following function is used:

$$c_i = a_7 s_i + a_8 d_i + a_9 r_i + a_{10} + z_i, \quad (11)$$

where c_i is the money cost on the i^{th} day, r_i is the average cut size on the i^{th} day, and z_i is a random variable with zero expected value.

7. Functions similar to those in Figures I and II are discussed by H. Simon, "A Behavioral Model of Rational Choice," this *Journal*, LXIX (1955).

8. Only the money costs incurred in the yard are included; the cost imputed to car delay is not taken into account. The categories of cost that are included in the data below are crew costs, clerical costs, costs of dispatching, oiling, inspecting, and switch-tending, administrative salaries, and fuel costs. Fixed charges are excluded, but many of these costs are fixed in the very short run. For some treatments of the cost of car delay, see E. Mansfield and H. Wein, "A Model for the Location of a Railroad Classification Yard," *Management Science* (1958), and "Linear Decision Rules and Freight Yard Operations," *Journal of Industrial Engineering* (1958).

9. However, some bias might arise in the regression coefficients estimated below if, for example, small actual outputs are associated with a negative discrepancy between actual and planned output and large actual outputs are associated with a positive discrepancy.

The effect of cut size on processing time and hence on costs has been noted above.

This cost function is not of the usual type that presumes factors are combined optimally for each output level. Our function is essentially an average relationship between cost and output. The basic data used below seem to indicate that a linear function is satisfactory, but no extensive tests were carried out. On a priori grounds, linearity does not seem unreasonable since these yards seldom operate near capacity.

Combining the cost equation (11) with the two output equations, (6) and (10), a three-equation model results:

$$\begin{aligned}s_i &= a_1b_i + a_2t_i + a_3 + u_i \\ d_i &= a_4A_i + a_5A_{i-1} + a_6 + v_i \\ c_i &= a_7s_i + a_8d_i + a_9r_i + a_{10} + z_i.\end{aligned}$$

(12)

The endogenous variables are s_i , d_i , and c_i . The remaining variables are exogenous. According to the model,

$$a_1 = a_2 = 1; \quad a_4, a_5 > 0; \quad a_4 + a_5 = 1.$$

(13)

VII. TESTS AND PREDICTION

To test the model, the a 's were estimated from data gathered at three freight yards. The estimates were then compared with the

TABLE I
SELECTED CHARACTERISTICS OF YARDS A, B, AND C, FALL 1955

Characteristic	Yard A	Yard B	Yard C
<i>Output</i>			
Number of cars switched per day	1,375	1,973	2,882
Number of cars delivered per day	485	261	149
<i>Input</i>			
Money cost per day	\$2,700	\$5,100	\$10,800
Engine hours used per day	141	264	379
<i>Other</i>			
Number of cars per cut	4.2	2.6	1.6
Backlog at end of day (average)	213	433	771
Number of classification tracks	22	47	70
Types of freight handled:	coal	autos	coal
	liquids	coal	steel
	steel	stock	autos

Source: Records of the co-operating railroad.

values in equation (13). To make these estimates, certain assumptions must be made about the distribution of the random variables

in equation (12). We assume that u_i , v_i , and z_i are normally distributed, homoscedastic, mutually independent, and serially independent.¹ Given these assumptions, it seems clear that least-squares estimates of the a 's will also be maximum-likelihood estimates.²

Note that the exogenous variables in the first two equations in (12) are forecasts. No data were available concerning the yardmaster's past forecasts and it was necessary to use instead the actual values corresponding to b_i , t_i , A_i , and A_{i-1} . We assume that the actual value equals the forecast plus a random variable with zero expected value and that the random variables are mutually independent, independent of the forecasts, and independent of the shock terms in the equations. Given these assumptions, the actual values may be substituted for the forecasts and the former may be treated as if they contained measurement errors.³

The three yards at which estimates of the a 's were prepared will be referred to as Yards A, B and C. Yards A and B are flat yards

TABLE II
LEAST-SQUARES ESTIMATES OF PARAMETERS AND STANDARD ERRORS
YARDS A, B, C

Parameters	Estimates and Standard Errors		
	Yard A	Yard B	Yard C
a_1	1.05(.07)	.98(.08)	.64(.07)
a_2	1.14(.03)	1.07(.05)	.72(.05)
a_3	-419	-560	316
a_4	1.04(.12)	.59(.16)	.78(.08)
a_5	.29(.11)	.05(.14)	.18(.07)
a_6	-25	86	8
a_7	.23(.08)	.14(.15)	.04(.16)
a_8	.21(.14)	2.67(.67)	.84(1.65)
a_9	-31.15(19)	-168.27(104)	431.04(855)
a_{10}	2370	5707	9791

Source: Basic data from records of co-operating railroad.

located in midwestern industrial centers; Yard C is a hump yard located in the Middle Atlantic states. Some relevant data concerning these yards are shown in Table I. At each yard, data for sixty-one

1. See footnote 4, below for the results of tests concerning some of these assumptions.

2. This can be seen by forming the joint probability density function of u_i , v_i , and z_i , by making a linear transformation to s_i , d_i , and c_i , and by forming the likelihood function, L . Maximizing L with respect to each of the parameters, the resulting estimators are those that result from the use of least squares.

3. Note that it is necessary for the variance of the "measurement errors" to be known. See K. Fox, "Structural Analysis and the Measurement of Demand for Farm Products," *Review of Economics and Statistics*, XXXVI (1954). See the following footnote for the results of some tests of the assumptions.

days were collected concerning the endogenous and exogenous variables, and least-squares estimates of the a 's were derived. The estimates were then adjusted for the "measurement errors," i.e., the differences between the forecasts and the actual values used.⁴ The resulting estimates are shown with their standard errors in Table II.

If the estimates are compared with the values in equation (13), the agreement is not too bad (considering the difficulties noted in the next paragraph). (1) The estimates of a_1 and a_2 are fairly close to one at Yards A and B; but at Yard C they depart from that value. (2) The estimates of a_4 and a_5 are positive at all three yards. (3) At Yard C, the sum of the estimates of a_4 and a_5 is fairly close to one. At Yards A and B, the sums differ considerably from one, but the standard errors are very large.⁵

In interpreting Table II, note that the estimates and standard errors depend on the variances of the "measurement errors." These variances could be estimated only roughly, and hence the estimates

4. To make these adjustments, estimates of the variances of the "measurement errors" are needed. Data were collected at Yard C concerning b_i and t_i and the corresponding actual values. The variances of the "measurement errors," or differences, were then estimated for Yard C. The corresponding variances for Yards A and B were estimated by assuming the ratio of the "measurement error" variation to the variation in the actual values to be constant from yard to yard. Similarly, data were collected at Yard B concerning the "measurement errors" for A_i and A_{i-1} ; and estimates of the variances at other yards were made in the same way.

Tests were carried out on these data to determine whether the "measurement errors" were mutually independent, independent of the random variables in the equations, and independent of the forecasts. The evidence did not call for the rejection of these hypotheses.

Tests were also carried out to determine whether after adjustment the residuals of the regression equations were mutually and serially independent. In general, the evidence did not call for rejection of these hypotheses, but some serial correlation seemed to show up in the cost equation residuals. These results are ambiguous, however, because the shock terms and the "measurement errors" are confounded in the residuals.

5. The coefficients in the cost equations seem to be consistent with the beliefs of most railroad men. (1) The three estimates of a_7 seem to be consistent with the belief that the marginal cost of switching is less at a hump yard (Yard A) than at flat yards (Yards B and C). (2) The interyard differences in the estimates of a_8 would be expected, since for cars that are delivered the average length of haul is greatest at Yard B and least at Yard A. Note, however, the large standard errors of the estimates.

Note that these costs include much more than the costs that are variable in the very short run and hence that the estimated cost equations do not contradict the first of the last three points in Section III.

Note too that the estimates of a_3 at Yard C and a_6 at Yards B and C are positive whereas we should expect them to be negative. The reason seems to be that the "measurement error" variances were underestimated in these cases and hence that the regression coefficients are underestimated and the intercepts are overestimated. Of course, sampling errors are present, too.

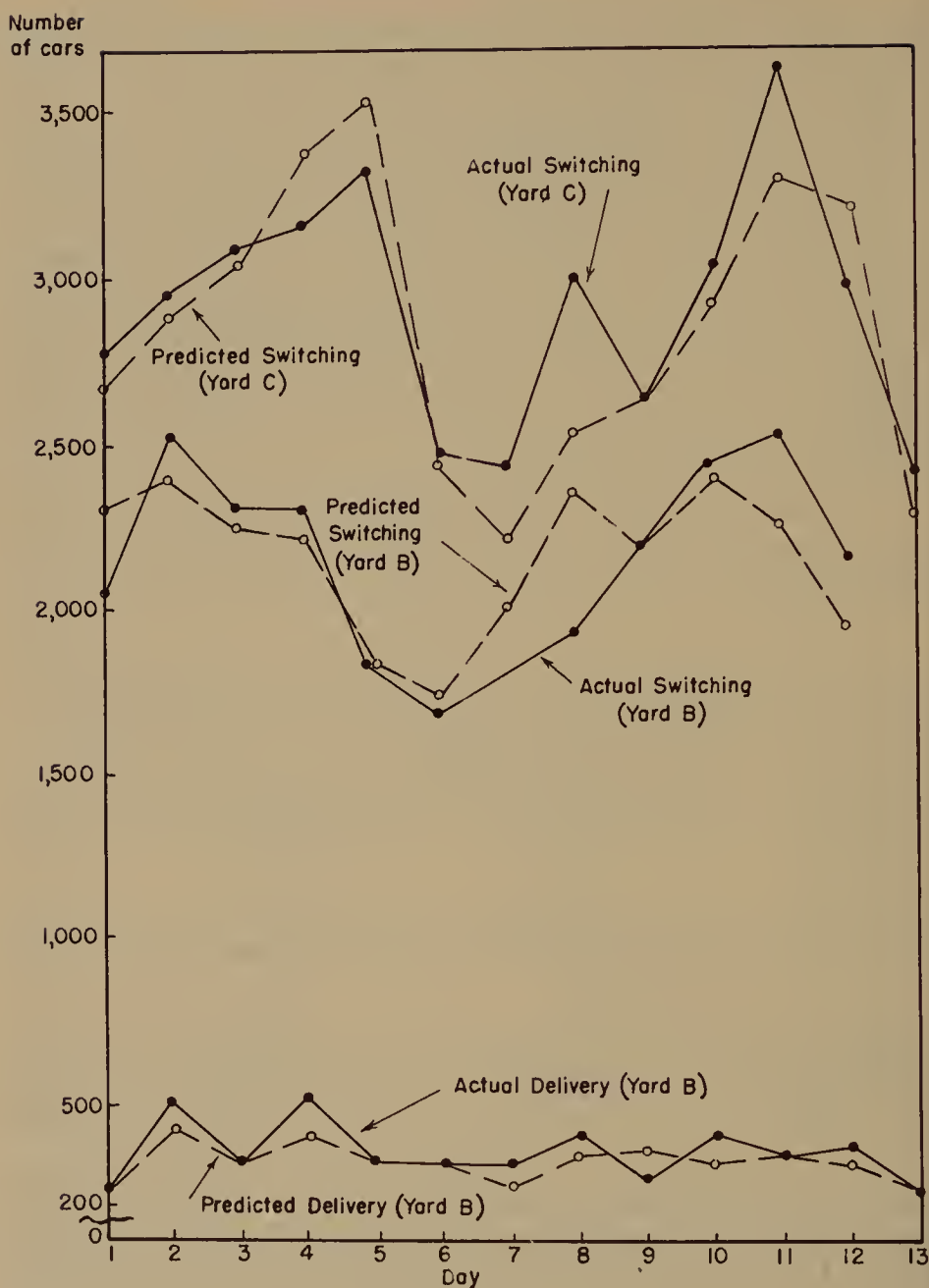


FIGURE III

SWITCHING AND DELIVERY PREDICTIONS, YARDS B AND C, 1956

and standard errors reflect errors other than those associated with sampling. These errors are difficult to estimate and they cloud the picture in Table II. The standard tests of the hypotheses in equation (13) would probably be misleading. The differences between hypothesized and estimated values should not be compared with the

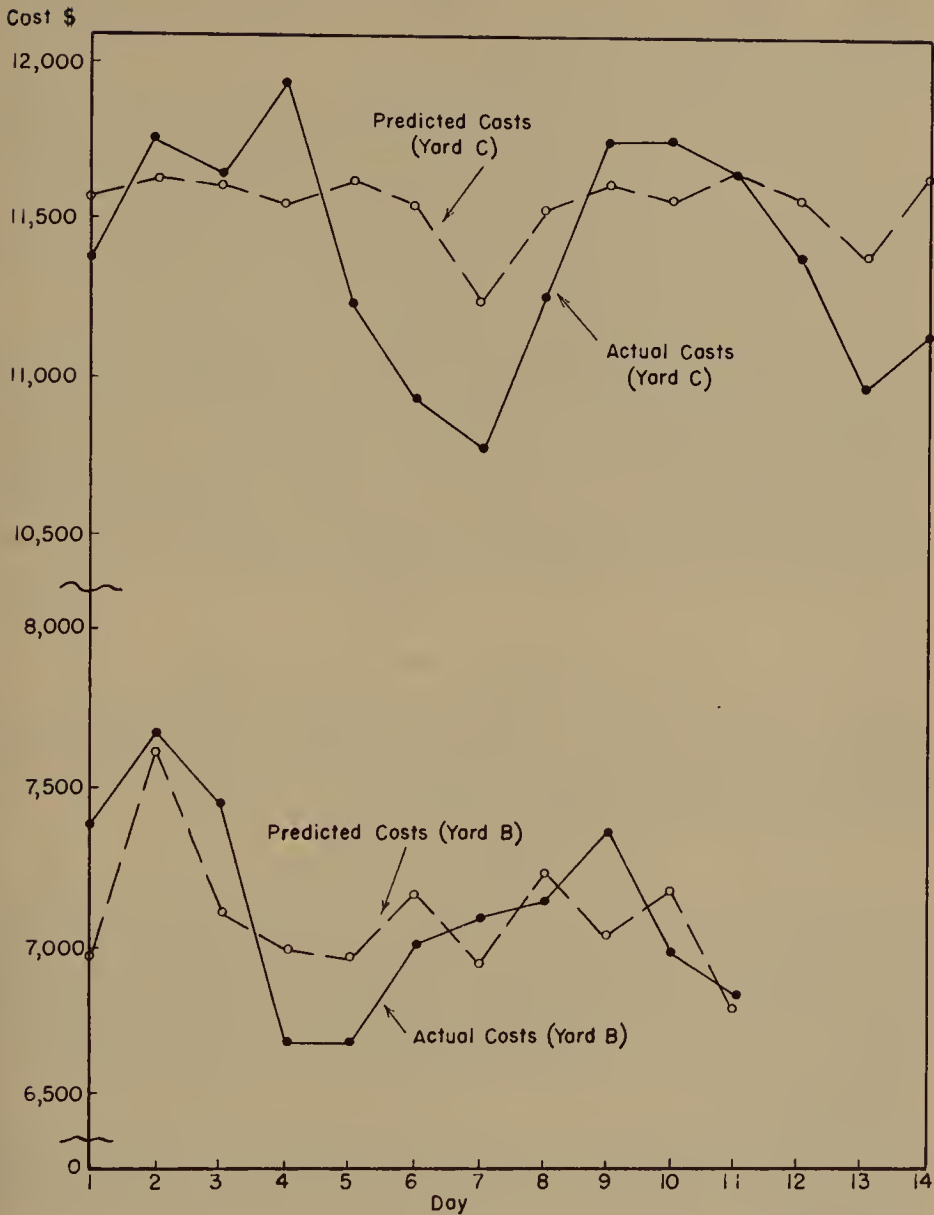


FIGURE IV
COST PREDICTIONS, YARDS B AND C, 1956

standard errors in the usual way that presumes only sampling errors are present.

To test the model further, an attempt was made to determine its predictive ability. For about two weeks, data were collected each day at Yards B and C concerning the values of the exogenous variables. In this case, the yardmaster's forecasts (b_i , t_i , A_i , A_{i-1}) could

be used. On the basis of this information and some of the estimates derived above, the switching output, the delivery output, and the costs were predicted for each day.⁶

The predicted number of cars switched and the predicted number of cars delivered are shown with the actual numbers in Figure III.⁷ The model predicts considerably better than the typical naive models.⁸ With regard to switching output, the model predicted the direction of change correctly in almost 90 per cent of the cases. With regard to delivery output, the results are somewhat less encouraging.

6. The period during which the predictions were made was almost a year subsequent to the sixty-one days referred to above. The predictions were based on the equations in (12). Of course, the shock terms shown there were suppressed. Both a_1 and a_2 were set equal to one, and a_3 was set equal to $\bar{s} - \bar{b} - \bar{t}$, where \bar{s} is the mean number of cars switched during the sixty-one day period, \bar{b} is the mean backlog, and \bar{t} is the mean number of cars that arrived to be switched. The value at each yard was quite close to that which results when rough estimates

are inserted in the right-hand side of $a_3 = \frac{\sigma^2 \log \alpha}{L_1 + L_2} + \frac{L_2 - L_1}{2} - F$.

In addition, a_4 was set equal to $\frac{\hat{a}_4}{\hat{a}_4 + \hat{a}_5}$; a_5 was set equal to $\frac{\hat{a}_5}{\hat{a}_4 + \hat{a}_5}$; and a_6 was set equal to $\bar{d} - a_4 \bar{A}_i - a_5 \bar{A}_{i-1}$; where \hat{a}_4 and \hat{a}_5 are the estimates in Table II, and \bar{d} , \bar{A}_i , and \bar{A}_{i-1} are the means of the actual values of d_i , A_i , and A_{i-1} for the sixty-one days. The values of a_6 are quite close to those that result when rough estimates are inserted in the right-hand side of

$$a_6 = \frac{\sigma'^2 \log \alpha'}{L_3 + L_4} + \frac{L_4 - L_3}{2}.$$

The estimates of a_7 , a_8 , a_9 , and a_{10} in Table II were used in the cost predictions. These predictions were then increased by 8 per cent because a wage increase had occurred since the sixty-one days. The cost predictions were made on the day following that to which the prediction applied. This was necessary because the value of r_i was not available until then. The actual delivery and switching outputs rather than those predicted by the model were used in the cost predictions.

7. Delivery output predictions are shown only for Yard B. At Yard C the yardmaster misunderstood our concept of delivery and the data were worthless. He included in his forecasts some types of cars that really did not have to be delivered.

8. The root mean square error of the model's predictions and the predictions of two naive models are shown below:

	Present Model	Naive Model ($x_i = x_{i-1}$)	Naive Model ($x_i = x_{i-1}^2/x_{i-2}$)
<i>Switching Output</i>			
Yard B	184.7	215.8	239.6
Yard C	210.3	459.3	701.8
<i>Delivery Output</i>			
Yard B	59.1	156.8	323.7
<i>Costs</i>			
Yard B	249.8	309.2	388.9
Yard C	333.6	351.9	471.8

The cost predictions in Figure IV are also superior to those based on naive models, though the results for Yard C are not much better than the latter. With regard to direction of change, the results are better than for delivery output but not so good as for switching output.

The model seems to stand up against these tests fairly well. But a final question might be considered. Does it predict better than a model that assumes the yardmaster minimizes expected costs? Our knowledge of the cost functions at Yards B and C is admittedly incomplete. But we used what data were available to set up a tentative cost-minimization model like that described in another paper.⁹ This model is an application of dynamic programming to the problem of output and employment determination. At Yard B it seems to predict switching output about as well as the model presented here. However, at Yard C it seems to be inferior. Note, however, that these results are based on sketchy information and that they are only tentative.¹

VIII. SUMMARY AND CONCLUSIONS

In this paper, a model is constructed to explain some decisions of a manager at an intermediate level of a railroad. These decisions are linked to actual fluctuations in output and cost, and the model is used to explain the observable fluctuations. The model is not based on profit maximization or cost minimization; a particular type of aversion function is used instead. Two types of tests are conducted: some coefficients prescribed by the model are compared with estimates, and the model is used for prediction. In both cases, the results seem quite encouraging.

The model is built around the manager's aversion functions. Behind these functions there lies the structure of penalties and incentive set up by his superiors.² To have much effect on his decisions, it would seem that changes in over-all company policy must be reflected in his aversion functions. His superiors attempt to influence these functions, presumably in accord with changes in over-all policy, by altering the structure of penalties and incentives.³ It might be interesting to study actual changes in over-all policy and to see, as best one can, the effects they have on the aversion functions and the decisions that are made. We conjecture that many changes in policy, by the time they trickle part way through a large organization, have surprisingly little effect on the latter.

9. Mansfield and Wein, "Linear Decision Rules and Freight Yard Operations," *op. cit.*

1. To minimize expected costs it appeared that the backlog planned at

The model also points to some problems involved in co-ordinating an organization in such a way that profit is maximized or some other such variable is optimized. Even if the optimum decisions were known, the penalty structure for each manager would have to be set so that these decisions would actually be made. In principle, it may be possible to tailor the penalty structure for each manager so that his decisions conform closely to the known optimum rules for production and employment scheduling, etc. But in practice, this may be very difficult because of the complexities involved in direct measurement of the aversion functions, the difficulties due to intertemporal instability in the functions, etc. Here, too, is an area where further study might be worthwhile.

Many important variables are under the immediate control of managers occupying intermediate positions in the firm. Yet, economics at present has little or nothing to say in this area. The present study is only a beginning: it deals with only one type of firm and only one level of the administrative structure within that firm. Hence, the results can be only tentative. But a model is presented that is sufficiently simple and operational to allow testing, and it seems to stand up to these tests fairly well.

CARNEGIE INSTITUTE OF TECHNOLOGY

Yard C should be about 500 cars. But the actual backlog hovers about 750 cars. Predictions based on a cost-minimization model overestimate the number of cars switched by about 250 cars. The larger backlog is quite compatible with the model presented here, one possible explanation being that the yardmaster's superiors are too much concerned with idle crew time. It is conceivable, of course, that it might also be compatible with a cost-minimization model if better data were available concerning costs.

2. The emphasis in preceding sections has been on penalties rather than incentives. Of course, incentives and penalties can both be easily incorporated into the model.

3. We do not mean that the "disutility" attaching to a particular penalty is altered, but rather that the penalty for a certain action (or event) is altered and hence that the "disutility" attaching to the action (or event) is changed. For example, $A(d)$ may be altered by changing the penalties attaching to particular discrepancies between the switching traffic and the planned output.

THE PLACE OF MONEY CAPITAL IN THE THEORY OF PRODUCTION

By ANDRÉ GABOR AND I. F. PEARCE

I. Outline of the problem of capital and of the need to reconcile its treatment by Austrian and neoclassical theorists, as also to establish a link between economics and accountancy practice, 537. — II. Introduction to the concept of the money capital requirement function and the criteria of profit rate maximization, 540. — III. Synthesis of divergent theories by full development of a new generalized money capital requirement function, 544. — IV. Comparison of the results here obtained with those of neoclassical theory, 555. — V. Summary, 557.

"Capital is not in the technical sense just another factor juxtaposed with the primary factors, but rather a new dimension on each of the primary factors." (Ragnar Frisch.)¹

I

The word "capital" is one commonly used by accountants, lawyers and economists in a bewildering variety of meanings, all loosely related to one another but by no means equivalent. In particular, capital is sometimes thought of as money, in Wicksell's words "the principal as opposed to the interest," and sometimes as durable equipment, i.e., the actual physical plant and buildings, required to carry on production.

Corresponding to these two points of view it is possible to distinguish two separate theories of production leading to two different concepts of profit. On the one hand, we have Wicksell and the Austrian school for whom capital is money and profit is indistinguishable from interest. Production is organized so as to maximize the rate of return on money invested.² On the other hand, what we shall call the neoclassical theory of production centers on the idea of a production function which includes money capital as a factor on the same basis as any other. Money can be freely purchased at a market price by the "entrepreneur" whose aim is to maximize his own remuneration which is the residual "profit."

Put this way there appears to be a more than trivial difference between the two schemes of thought, although they purport to

1. "Frisch on Wicksell," *The Development of Economic Thought — Great Economists in Perspective*, ed. H. W. Spiegel (New York: Wiley and Sons, 1952), pp. 653-99.

2. Cf. Knut Wicksell, *Lectures on Political Economy* (English ed.; London: George Routledge and Sons, 1934), I, 181, equation (i).

describe the same empirical phenomena. And yet, it is generally accepted that both arguments amount in the end to the same thing. It seems to be agreed that the full content of the Austrian theory can be imported into the neoclassical framework simply by qualifying the marginal value products of static theory with the adjective "time discounted."³ The important distinction between capital (fixed equipment) which contributes to production, and capital (work in progress) which is itself part of the product, tends now to be forgotten. Wicksell's conception of money as the controlling factor, in the interest of which the firm operates, has been abandoned in favor of the "entrepreneur."

It is not difficult to find very good reasons for the relative eclipse of the Austrian approach. In the first place, it has never been successfully developed with any very great generality. In its theoretical expositions, production almost always consists of wine maturing in a vacuum or trees growing upon land of no value.⁴ Again it can be argued, and indeed Wicksell himself showed, that whichever theory of profit we choose, the *general* equilibrium attained will have the same properties. We know, in fact, by the so-called "adding up" criterion, that if all factors except one are paid at the rate of the value of their marginal product, the residual will always be the marginal value product of the excluded factor, *as long as production takes place at minimum average cost and price equals average cost.*⁵ Thus general equilibrium is the same whether management is maximizing its own reward while paying money capital the value of its marginal product, or vice versa.

On the other hand, there is much to suggest that a great deal has been lost by the failure to produce a more adequate synthesis of all that is best in the work of both the Austrian and the neoclassical schools. In the first place, the fact that two theories of profit lead to the same general equilibrium is not sufficient to make them the same theory. The route by which equilibrium is attained is often as important as the equilibrium itself. We have shown elsewhere⁶ that, if general equilibrium has not been attained, and the fact that an investment is being contemplated in any industry implies that it has not, then the two theories of profit, the Wicksellian and the neo-

3. Cf. Earl Rolph, "The Discounted Marginal Productivity Doctrine," *Journal of Political Economy*, XLVII (Aug. 1939), 542-56.

4. Cf., e.g., Friedrich and Vera Lutz, *The Theory of Investment of the Firm* (Princeton University Press, 1951).

5. Cf. Paul Samuelson, *Foundations of Economic Analysis*, p. 86.

6. André Gabor and I. F. Pearce, "A New Approach to the Theory of the Firm," *Oxford Economic Papers* (New Series), LIV (Oct. 1952), 252-65.

classical, lead to different conclusions. Both optimum scale and optimum technique of production will be different in each case. This follows from the fact that, unless all producers are operating at minimum average cost, the "adding up" criterion no longer holds. We come back to this again later from a different point of view.

Nor is this all. There is no place anywhere in the neoclassical theory of the firm for working capital. It cannot logically be included in the production function since at least some part of it must be the money required to sustain other factors throughout the period of production; and this will vary with the prices of those factors. Thus, the *quantity* of one factor depends upon the *prices*, of the others. In other words, the production function is a function of factor prices as well as factor inputs. This makes nonsense of the apparatus of equi-product contours and constant outlay lines in terms of which the neoclassical theory is commonly set out.⁷

Again, the student of the modern textbook theory of the firm, who is also a student of accountancy, will find very few points of contact between his two subjects. Profit in one is a cost in the other and vice versa. Neoclassical theory does not deal explicitly with the role of money whereas this is the beginning and the end of the accountant's interest. The assumption of an entrepreneur, who can scarcely be identified in practice, brushes aside all questions of who is acting in whose interest, and with it all need to consider the merits of ordinary versus preference shares.⁸ The neoclassical firm comes into existence with the wave of a wand. There is no pre-production period which the accountant can identify as corresponding to his ritual of share issues and calls. The life of durable equipment does not figure in the theory, so that the discussion of depreciation funds appears as an uncomfortable appendage only reluctantly introduced, if it is introduced at all.

We believe that in the present paper we have succeeded in a modest degree in maintaining a footing in all three camps, the accountants', the Austrian and the neoclassical. A scheme is proposed which has the neoclassical advantage of a high level of generality, at the same time retaining a build-up period which serves to reveal explicitly the nature of capital and its relationship to the accountant's concept. Our approach in fact makes it necessary to seek the theoretical solution of a problem which has long been a source of trouble to accountants and lawyers, namely that of dis-

7. Cf. Joan Robinson, *The Accumulation of Capital* (London: Macmillan, 1956), p. 414 footnote.

8. Cf. Gabor and Pearce, *op. cit.*

tinguishing between the payment of profits and the repayment of capital. The solution obtained enables us to generalize and give a nontautological meaning to Wicksell's formula for capital recently rediscovered by Professors R. Kahn and D. G. Champernowne.⁹ It becomes possible also to compare and contrast in a convenient way the two theories of profits referred to above. In the first instance we accept the Wicksellian hypothesis. Our firm is controlled by ordinary shareholders who buy "management" (which is regarded as a form of labor) at its market price. Management operates so as to maximize the rate of return on ordinary shares. Later we show how the neoclassical assumption of an entrepreneur fits into our scheme.¹ We make no secret, however, of our preference for Wicksell's view which we believe to accord more closely with the facts and which leads naturally to issues which are important in the real world. In fact, the present paper represents in part a plea for the return to the idea of a firm as an investment opportunity with management acting on behalf of the suppliers of capital.

Economists who do not share our views, tend to argue that loan capital is in fact money bought in the market at a fixed price, exactly as assumed by the neoclassical theorists. But all shares cannot be preference shares, and it is easy to show that if preference and ordinary shares are taken in a fixed ratio (the so-called gearing ratio), the equilibrium output is precisely the same as if all shares were ordinary. And even if there were no fixed gearing ratio, it would still be more profitable to any limited number of ordinary shareholders to use borrowed funds to become ordinary shareholders in a second firm rather than to extend their own firm into the region of decreasing returns.²

II

We begin from the idea of a production function which we shall write

$$F = F(x, y, \dots),$$

where F is the output per unit of time of the product or service, and x, y , etc., the inputs of factors per unit of time, *none of the x, y , etc., being money capital*. The exclusion of money capital from the production function is the cornerstone of our approach. It seems clear to us that there is an essential difference between money and any

9. Cf. Joan Robinson, *op. cit.*, final appendix.

1. Cf. especially Section IV below.

2. Cf. Gabor and Pearce, *op. cit.*

other factor. Production could conceivably be carried on without money if other inputs were made available by some dictatorial hand; but the removal of a unit of any other factor would immediately reduce the product. Money capital is a catalyst. It has no direct marginal product but operates only by influencing the way in which other factors are used. We return to this point again towards the end of Section III.³

The inputs x, y , etc., may be single use goods, e.g., raw materials, or durable machines, quantities being measured in terms of appropriate physical units. It should be noted that if x in the production function stands for x machines we do not mean that x machines were *used up* in the time period. We mean only that they were *in use*. The fact that single use goods are used up in the time period but durable machines are not is irrelevant as far as the writing of the production function is concerned. It becomes relevant only later when we attach a life span to each factor.

Stocks of raw materials and work in progress are not factors of production but they do influence the money capital requirement in a way which will become evident later. In brief, our production function corresponds exactly to the production function which figures in modern theories of growth. Capital accumulation means the accumulation of durable factors of production in our sense. The accumulation of stocks and work in progress does not contribute directly to production but affects it indirectly via the rate of interest.

In addition to the production function, we have a money capital requirement equation.⁴ This expresses the dependence of the money capital requirement on the prices and quantities of all other factors used. Most of the next section will be concerned with the investigation of the exact nature of this dependence. In the meantime, however, we assume factor prices to be given and write money capital requirements as

$$C = C(x, y \dots),$$

making C a function of all other factor quantities only. C may be made up of ordinary shares or of borrowed funds or of any mixture of both. But, as earlier claimed, we lose nothing at this stage by assuming the whole to be ordinary capital. For convenience therefore we introduce this hypothesis.

The rate of return r on ordinary shares therefore is, by defini-

3. Cf. also Oskar Lange, "The Place of Interest in the Theory of Production," *Review of Economic Studies*, III (June 1936), 159-92.

4. *Ibid.*

tion, total revenue minus total cost divided by the money capital requirement, i.e., in the two factor case,

$$r = \frac{mF - px - qy}{C}, \quad (\text{i})$$

where m , p and q are the prices of the product and of factors x and y , respectively. (The precise meaning of the price if the factor is durable equipment will be discussed further down.)

We now consider the choice of the best technique which resolves itself into the problem of finding the values of x and y which will maximize r for any given level of output, i.e., subject to the constraint that F equals a constant. In other words, we want to find that point on any given equi-product contour where the rate of return on capital is at a maximum. Since at that point the marginal rate of return equals the average rate of return,⁵ we obtain

$$\frac{qF_x - pF_y}{C_xF_y - C_yF_x} = \frac{mF - px - qy}{C} \quad (\text{ii})$$

as the maximum condition. In view of (i), this implies that for equilibrium values of x and y (m , p and q constant)

$$r_{max} = \frac{qF_x - pF_y}{C_xF_y - C_yF_x}.$$

This can also be written in the form

$$\frac{F_x}{F_y} = \frac{p + r_{max}C_x}{q + r_{max}C_y}. \quad (\text{iii})$$

In economic terms, equation (iii) means that at every possible level of output the factors of production will be so combined that the ratio of the marginal physical productivities is equal to the ratio of the cost of using a unit of each factor per unit of time, including an appropriate charge for capital.

The point we wish to emphasize, however, is not that a charge for capital must be made, for this would be everywhere agreed, but that the charge must be calculated at a rate equal to the maximum rate it is possible to earn in the projected investment at each level of output. r_{max} will not be the same for every level of output. Nor will it be the same in general as the objectively given market rate of interest, even for that level of output at which r_{max} is greatest.

5. Cf. Lutz, *op. cit.*; also Tibor Scitovsky, "Maximation du profit," *Revue Economique*, No. 3 (May 1955), pp. 368-86.

It will be the same only at the point where both the firm and the industry are in equilibrium.⁶ It depends on conditions which are not under review here whether, in any particular case, such a point exists or not. The possibility of several points or even a region of such double equilibrium cannot be ruled out either.

If we had begun from the assumption that the rate of interest is always given in the market and that there exists in the firm an "entrepreneur" whose aim is to maximize his residual income, we should have reached a conclusion which looks formally the same as equation (iii) above. The difference would be that in the place of r_{max} we should have r_o , the given market rate. If $r_o \neq r_{max}$, this would imply a different technique of production.

It is worth noting in parenthesis that the distinction between r_o and r_{max} is very relevant to the celebrated dispute over the so-called "Ricardo effect," i.e., over the extent to which changes in the profit margin or the market rate of interest will affect the degree of capitalization of productive processes.⁷ Only if general equilibrium continuously prevailed should we have r_o and r_{max} identical. In macro-economic arguments it might sometimes be convenient to assume that, when there is change, r_o and r_{max} change together, but if so, the assumption should be made explicit. It should be emphasized that any change is from one position of general equilibrium to another.

On the other hand, out of equilibrium situations are very relevant to the individual firm's decisions, and in this paper it is with individual firms' decisions that we are primarily concerned. In what follows we shall need to maintain carefully the distinction which we have drawn between r_o and r_{max} .

Consider now the optimum *scale* of output as opposed to technique. Equation (iii) can be written

$$\lambda F_x = p + r_{max}C_x, \quad \lambda F_y = q + r_{max}C_y, \quad (\text{iv})$$

where λ is an undetermined multiplier, the same for all factors. If $\lambda < m$, the price of the product, then it will pay to increase the scale of output until $\lambda = m$. It will be remembered that r_{max} is the greatest rate of return for any *given* level of output. In the special case where $\lambda = m$, r_{max} is a *maximum maximorum*, i.e., it is the greatest rate of return it is possible to earn at any level of output.

In this case

$$mF_x = p + r_{max}C_x, \quad mF_y = q + r_{max}C_y. \quad (\text{v})$$

6. Cf. Section IV below.

7. Cf. Nicholas Kaldor, "Professor Hayek and the Concertina Effect," *Economica*, Vol. 9 (Nov. 1942), pp. 359-82.

This may be compared with the proposition familiar in the neo-classical theory of production

$$mF_x = p, \quad mF_y = q, \quad (\text{vi})$$

i.e., the price of any factor in equilibrium is equal to the value of its marginal product.

The results (v) and (vi) can, of course, be reconciled by defining p and q in (vi) so as to include the appropriate charges for capital, $r_{max}C_x$ and $r_{max}C_y$, respectively. If x is machinery, most economists would readily accept the necessity for including this profit charge in one form or another. It is less certain, however, whether the need to include interest (profit) would be as widely recognized if x were labor. Moreover, the neoclassical theory gives no hint of any difficulty regarding the proper profit *rate* to charge; nor is it at all clear what is meant by the price p when x is a durable use good with a life greater than the time period. In practice, the total cost $px + qy$ will not be paid out evenly in each period of time. Durable factors are purchased only as and when old plants wear out. To make sense of equation (iii) it is necessary to be able to attach a precise meaning to the p , q , C_x and C_y which appear in it. It is to this problem that we now turn.

III

The need for money capital exists only because in practice a flow of production cannot be instantaneously created. There must always be a build-up period which is all important in the determination of equilibrium and even of the proportions in which factors are to be used.

We assume that the aim is to set up a continuous even flow of product. In order to create this flow at any given moment of time it will be necessary to purchase and employ factors *before* that time. The flow of any input once commenced is assumed to be continuous. Where the input is durable equipment it will be used continuously from the moment of its first purchase until the end of its life l when it will at once be replaced. The lapse of time between the payment for the first unit of a factor and the commencement of the product flow which it is designed to create, (possibly in combination with other factors,) will be called the *maturity time* t of that factor. It will be seen that the maturity time is a new and vitally important dimension of every factor. Negative maturity times are not ruled out; labor engaged in repairs to durable equipment may, for example, have a negative maturity time. It will be noted that it is essential

to distinguish factors according to maturity time as well as function.⁸ In our model x and y may both be labor, homogeneous in the sense that they are doing the same job, but if they are employed at different points in the productive process, they will have to be defined as different because they have different maturity times.

Now the money capital requirement of the firm at any moment of time is simply the difference between its total receipts and payments up to that time. We therefore proceed to count up expenditure. Suppose that the flow of output has been established for T units of time. The first x_i units of the i^{th} input will have been purchased $T + t_i$ units of time ago, since t_i is by definition the maturity time of factor i . Total expenditure, excluding profits, on the full set of first units is

$$\sum_{i=1}^n p_i x_i .$$

It will simplify the analysis if we assume that profits are paid to the owners of capital from the date of the investment. (Such a practice, though not usual, is not by any means unknown. In any case, it would not materially alter the results if profits not covered by current net revenue were added to capital.) Total expenditure and profit on the first set is therefore

$$\sum_{i=1}^n p_i x_i (1 + r)^{T + t_i} .$$

The second x_i units of the i^{th} factor will have been purchased l_i units of time after the first, since the i^{th} factor has a life of l_i . Profits will have been paid on this for $T + t_i - l_i$ time units. Expenditure to date on the full requirement of second units of all factors will therefore be

$$\sum_{i=1}^n p_i x_i (1 + r)^{T + t_i - l_i} .$$

For the set of third units it will be

$$\sum_{i=1}^n p_i x_i (1 + r)^{T + t_i - 2l_i} ,$$

8. Cf. Arthur Smithies, "The Austrian Theory of Capital in Relation to Partial Equilibrium Theory," this *Journal*, L (Nov. 1935), p. 125.

and so on. The number of times $(S_i + 1)$, it will have been necessary to purchase x_i units of i up to time T will be the number of lives l_i in $T + t_i$ units of time plus one. Hence we define

$$S_i + 1 = \frac{T + t_i - h_i}{l_i} + 1,$$

where h_i/l_i is that fraction of l_i by which $(T + t_i)/l_i$ exceeds its nearest whole number.

The grand total of all expenditure to date is given by the expression

$$\sum_{i=1}^n \sum_{s=0}^{S_i} p_i x_i (1 + r)^{T + t_i - s l_i}.$$

So far we have charged profit on all expenditure, which implies that all expenditure is out of capital. The firm will not, of course, be required to pay profits on any expenditure made out of revenue. To give expression to this, we deduct from total expenditure not only revenue, but revenue compounded at the profit rate from the moment of receipt to the present, i.e., we deduct

$$\sum_{t=0}^T mF(1 + r)^t.$$

The balance must be capital put up by shareholders. Hence we have

$$C = \sum_{i=1}^n \sum_{s=0}^{S_i} p_i x_i (1 + r)^{T + t_i - s l_i} - \sum_{t=0}^T mF(1 + r)^t. \quad (\text{vii})$$

Three things should be noted about this expression for C . Firstly, any sums of money put aside out of revenue to build up a depreciation fund are excluded from the definition of the capital tied up. This is logical since depreciation funds are in any case not "tied up," or, if they are, they must be tied up in some investment other than that against which they are charged.

Secondly, we now see that C is a function of r , the rate of profit paid out. Earlier, when we obtained our condition for r_{max} ,⁹ we found it convenient to imply, without so stating, that C was in fact independent of r . It is now clear that, in setting down the change in total capital requirement due to changes in factors x and y (as included in equation (ii)), we ought strictly to have added another

9. Cf. equation (ii) above.

term representing the change in C due to the changed profit rate. Fortunately, by definition, the maximum profit rate r_{max} is that value of r at which its rate of change is zero. Hence the value of the extra term would in any case have been zero in the region of equilibrium. Within the limits of the restriction so imposed, it would be a simple matter to go back and derive our basic result (iii) quite rigorously.

Thirdly, it is a particularly satisfactory feature of our approach that it has now brought us to another difficult accountancy problem in such a way as to suggest a theoretical solution.

Our original definition (i) of r , the rate of profit, is meaningless until we have defined C . Our definition of C in (vii) is meaningless until we have defined r . We have not yet stated a way of distinguishing on the one hand between profit payments and capital repayments, and, on the other, between expenditure out of revenue and expenditure out of profits withheld from shareholders. This is a familiar problem which has always been a source of trouble to accountants. Indeed, a not inconsiderable body of case law has been built up around the question, mostly based on value judgments of one kind or another. It is a remarkable fact, however, that a very weak condition which will command ready acceptance, is quite sufficient to enable us to define profits precisely for a given continuous process.

It is clear that C in equation (vii) is a function of time T . It is also clear that if r is set too high, C will become very large when T is large. Shareholders' profits can then be paid only out of new issues of capital. If r is set too low, the firm will be building up depreciation funds or repaying capital at such a rate that C will become large and negative when T is large.

In order to define r , therefore, we make the weak assumption that it is chosen so that, *however long the process is carried on*, C will never become very large and positive or very large and negative. In other words, we require C to fluctuate indefinitely around some kind of average as long as there is no change in the technical process carried on. Let us now see what this implies.

Carrying out the summations in (vii) and manipulating we obtain

$$C(1+r)^{-T} = \sum_{i=1}^n p_i x_i (1+r)^{t_i} \frac{1 - (1+r)^{-(T+t_i+z_i)}}{1 - (1+r)^{-l_i}} + mF \frac{(1+r)^{-T} - (1+r)}{r}, \quad (\text{viii a})$$

where $z_i = l_i - h_i$.

Equation (viii a) may be rewritten

$$\sum_{i=1}^n p_i x_i \frac{(1+r)^{t_i}}{1 - (1+r)^{-l_i}} - \frac{mF(1+r)}{r} = (1+r)^{-T} \left[C - \frac{mF}{r} + \sum_{i=1}^n \frac{p_i x_i (1+r)^{-z_i}}{1 - (1+r)^{-l_i}} \right], \quad (\text{viii b})$$

which can be seen to be of the form

$$A = (1+r)^{-T} B, \quad (\text{viii c})$$

where both A and B are independent of T . We require (viii c) to be true for all T , and this is only possible if $A \equiv 0$, $B \equiv 0$, i.e., if

$$mF = r \sum_{i=1}^n p_i x_i \frac{(1+r)^{t_i-1}}{1 - (1+r)^{-l_i}}, \quad (\text{ix})$$

and

$$C = \frac{mF}{r} - \sum_{i=1}^n p_i x_i \frac{(1+r)^{-z_i}}{1 - (1+r)^{-l_i}}. \quad (\text{x})$$

Equation (ix) gives us the definition of profit in terms of revenue flow, costs, equipment lives and maturity times which we set out to obtain.

Substituting for mF from (ix) into (x) we get

$$C = \sum_{i=1}^n p_i x_i \frac{(1+r)^{t_i-1} - (1+r)^{-z_i}}{1 - (1+r)^{-l_i}}. \quad (\text{xi})$$

It is easy to see that as long as the variables in equation (xi) are positive and do not assume erratic values which are of no relevance to the real problem at issue, our condition that C should never increase to infinity is satisfied.

This formula is basic. We shall see below how all of the well-known results of capital and interest theory are special cases of it. It is completely general. It holds whether the contemplated investment is an equilibrium one or not, and whether the industry is in equilibrium or not. r need not be r_{max} , nor need it be the same as the given market rate. It is defined simply by (ix).

From (xi), given technical details and factor prices, capital requirement can be computed for any point of time.

Although time T does not appear explicitly in (xi), capital requirement in the general case still varies over time. This is because all z_i are functions of time; by definition, z_i is the difference between $(T + t_i)$ and the lowest number greater than $(T + t_i)$ which is divisible by l_i without remainder. Over time, therefore, z_i takes values $1, 2, 3, \dots, l_i, 1, 2, \dots$. The i^{th} term in the summation (xi) is at a maximum when z_i is equal to l_i , and at a minimum when $z_i = 1$. If the paid-up share capital of our firm is fixed, its reserve fund will be the difference between this and C . Hence the fluctuations in C are observable in practice as fluctuations in depreciation reserves.

In general, when the product flow commences, C will be greater than the mean value of C (which we shall write \bar{C}); subsequently C will fluctuate round \bar{C} . If the various items of equipment have different lives, this will tend to depress the mean deviation of C from its average value. If all equipment were in a balanced age composition, C would be constant over time and equal to \bar{C} . It is probably true that if the number of inputs is large, big divergencies between C and \bar{C} will be rare. For this reason we propose to find an expression for \bar{C} , and treat C as if it were a constant at \bar{C} . This will enable us to reinstate the static picture, and it is hoped that it will not do too great injustice to the facts. (It has been shown in an unpublished paper by Professor T. W. Swan and Mr. E. J. Hannan that, if all amortization charges were ploughed back, any initial investment at a given moment of time is bound to spread eventually to a balanced age composition of equal value.)

From the form of (xi) it is clear that to find \bar{C} we need only find the average value of all $(1 + r)^{-z_i}$ over the period l_i . This is given by

$$\frac{1}{l_i} \sum_{z=1}^{l_i} (1 + r)^{-z} = \frac{1 - (1 + r)^{-l_i}}{rl_i},$$

hence

$$\bar{C} = \sum_{i=1}^n p_i x_i \left[\frac{(1 + r)^{t_i - 1}}{1 - (1 + r)^{-l_i}} - \frac{1}{rl_i} \right]. \quad (\text{xiii})$$

If $t_i = 1$ and $l_i = 1$, (xii) reduces to

$$\bar{C} = \sum_{i=1}^n p_i x_i.$$

This is the form of capital function which Lange adopts.¹ If $t_i = 1$ and the time period is very short so that we may substitute the

1. *Op. cit.*, p. 174.

exponential expression for the corresponding continuous rate of interest in our formulae wherever we now have $(1 + r)$, then (xii) takes the *same form* as Wicksell's expression for the *value* of capital equipment.² This has also been referred to as the Kahn-Champernowne formula.³ Again, if all goods are single use goods, so that $l_i = 1$, then (xii) becomes

$$\bar{C} = \sum_{i=1}^n p_i x_i \frac{(1 + r)^{t_i - 1}}{r},$$

which is Wicksell's formula for the *value* of social capital in the wine case.⁴

It is of the utmost importance however to recognize that although (xii) takes the same form as the general expression for the value of social capital, it is most emphatically *not* the same thing. Equation (xii) is the *physical amount of money required to sustain the productive process defined by the set of inputs x_i , whether the process is an economic optimum or not*. There is no implication of equilibrium. It is not surprising, of course, that the money capital requirement is equal to the value of work in progress when the valuation is made on the basis of the *average* earnings of money in the process. The same is *not* true in general if the valuation is made on the basis of the *marginal* earnings of money in the process. Nor is it true, in general, that the marginal increase in capital requirement due to unit increase in factor x_i (i.e., \bar{C}_i), is equal to the value of the increase in capital calculated on the basis of marginal earnings, for, differentiating (xii) with respect to x_i , we have

$$\bar{C}_i = p_i \left[\frac{(1 + r)^{t_i - 1}}{1 - (1 + r)^{-l_i}} - \frac{1}{r l_i} \right] + \bar{C}_r r_i, \quad (\text{xiii})$$

where \bar{C}_r is the change in \bar{C} due to unit change in r , and r_i is the change in r due to unit change in x_i . The extra term $\bar{C}_r r_i$ is the analogue for the individual firm, as opposed to the economy as a whole, of the celebrated "Wicksell effect." It is the revaluation of capital due to the change in the rate of profit r . Only in the case where r_i is zero, (i.e., when $r = r_{max}$), will this effect be zero.

To define capital as the value of equipment and work in progress involves a subjective element and all the difficulties attendant upon it. As in the case of consumption goods, the value of capital depends

2. *Op. cit.*, p. 283, equation (15) with (4), p. 276, substituted.

3. Joan Robinson, *op. cit.*, final appendix.

4. *Op. cit.*, p. 179, equation (4).

on how much of it you have. The money required to produce, however, is objectively determined by the prices of inputs and product and by the technical data, i.e., it is given by equations (ix) and (xii) together. There is no ambiguity about the meaning of r which is defined by (ix). If, on the other hand, we say with Kahn and Champernowne that our capital is the sum of the discounted future earnings of all the things we have, we may be begging the question of the appropriate rate at which we should discount. Should it be the marginal internal rate of profit or the average rate? Or should it be the average market rate? If we are interested in problems of capital accumulation, might not the appropriate rate be the marginal rate of return on an increase in national capital? It is clearly question begging to define the equilibrium of the firm as the point where the value of the marginal discounted earnings of each input is equal to its price, without at the same time defining the appropriate discounting rate; for it is always possible to choose some rate where this is true at *every* level of output. In our account of the theory of the firm, equilibrium is determined by equation (v) above. We now proceed to see what this implies.

From (x) we can obtain a further identity:

$$rC \equiv mF - \sum_{i=1}^n p_i x_i r \frac{(1+r)^{-z_i}}{1 - (1+r)^{-l_i}}.$$

At first sight this looks like a contradiction, since it appears to suggest that profit is something different from revenue minus cost. But we recall that C in fact fluctuates over time. The factors multiplying each $p_i x_i$ fluctuate over time correspondingly so as to maintain the identity. If the identity holds for all points of time, it holds also for the average values of both sides. Deriving the mean of the right hand side as before (cf. equation (xii) above) we can write

$$r = \frac{mF - \sum_{i=1}^n (p_i x_i / l_i)}{\bar{C}}.$$

Comparing this with equation (i) above, (the definition of r with which we began),

$$r = \frac{mF - px - qy}{C}, \quad (i)$$

it becomes apparent that if we wish to keep to the static picture we must replace C in Section II with \bar{C} . We also see that if p and q

are the market prices of physical units of inputs x and y respectively, we should have written p/l_x and q/l_y rather than p and q . This corresponds to the notion of inputs as a flow. We now have what we set out to obtain, namely, an exact definition of the price of an input and of C and hence of \bar{C}_x .

Our equilibrium condition (iii) above must now be written

$$\frac{F_x}{F_y} = \frac{p/l_x + r_{max}\bar{C}_x}{q/l_y + r_{max}\bar{C}_y}.$$

Substituting into this the expressions for \bar{C}_x and \bar{C}_y obtained from (xiii), and remembering that, if r is to be r_{max} , the terms $\bar{C}_x r_x$ and $\bar{C}_y r_y$ must be proportional to F_x and F_y , and can therefore be neglected, we obtain

$$\frac{F_x}{F_y} = \frac{p \frac{(1+r)^{t_x-1}}{1 - (1+r)^{-l_x}}}{q \frac{(1+r)^{t_y-1}}{1 - (1+r)^{-l_y}}}, \quad \text{where } r = r_{max}$$

which may be written

$$p = \lambda \frac{mF_x(1+r)^{-(t_x-1)}}{r} \left[1 - (1+r)^{-l_x} \right],$$

or

$$p = \lambda \sum_{t=t_x}^{(t_x+l_x-1)} mF_x(1+r)^{-t}, \text{ and analogously for } q. \quad (\text{xiv})$$

In words, factor prices must be *proportional* to the discounted value of their future earnings, the discount rate being the maximum rate it is possible to earn at the given level of output. If r is a *maximum maximorum*, the proportionality becomes equality.

Our basic formula for equilibrium is revealed as an obvious generalization of Wickcell's.⁵ If the input is a single use good, so that $l_x = 1$, we have

$$p(1+r)^{t_x} = mF_x, \dots (r = r_{max}), \quad (\text{xv})$$

which, if the time period is small, is precisely Wickcell's conclusion.

Finally, we draw attention to the facts:

(a) that the formulae for the "real" rate of interest given by Jevons, Wickcell, J. B. Clark and Lange⁶ are all special cases of our

5. *Op. cit.*, p. 181, equations (2) and (3).

6. Cf. Lange, *op. cit.*, p. 169.

maximizing condition $mF_x = p/l_x + r_{max}\bar{C}_x$;

(b) that all these formulae depend for their validity on the assumption of equilibrium;

(c) that in equilibrium they are all special cases of the obvious statement that the rate of interest is equal to profit divided by capital tied up, i.e., as in our formula

$$r = \frac{mF_x - px/l_x - qy/l_y}{\bar{C}}.$$

It is a weakness of Wicksell's wine case that it implies that in some sense the existence of a "real" rate of interest depends upon "time" being productive. That is, it depends on the fact that wine improves with age. *The real rate of interest exists because production takes time, and not because time is productive.* We shall illustrate the distinction by consideration of a case where a given input is employed on two kinds of jobs each with a different maturity time. Here time is clearly nonproductive. The marginal products of the two kinds of input are different because equilibrium demands that they should be used in proportions which make them different, and not because their product matures in any sense with the passage of time. Suppose x has a maturity time t , and y a maturity time $t + 1$, then from equation (xiii) we have, in equilibrium,

$$\bar{C}_y - \bar{C}_x = \frac{p(1+r)^t(1-(1+r)^{-1})}{1-(1+r)^{-t}}.$$

(p and l are, of course, the same for factors x and y , since by definition x and y are physically identical.)

Assuming with all of the writers mentioned above that inputs are single use (i.e., $l = 1$), this reduces to

$$\bar{C}_y - \bar{C}_x = p(1+r)^t.$$

Now our equilibrium condition, marginal rate of return to equal average, can be stated quite generally in words. We require r to be equal to the change in total profits, arising out of any change in technique or scale, divided by the corresponding change in money capital requirement. Consider the substitution of a unit of factor y for a unit of factor x in the region of this equilibrium. There will be no change in total cost, since x and y have the same price. The change in total revenue will be $m(F_y - F_x)$, whilst the change in capital requirement is $p(1+r)^t$ as shown above, hence

$$r = \frac{m(F_y - F_x)}{p(1+r)^t}. \quad (\text{xvi})$$

This is equivalent to all of the real rate of interest formulae referred to on the special assumptions appropriate to each case. Wicksell's wine model supposes production to be carried on with one unit of product (one hectolitre of grape juice maturing to time t). We may suppose this to be factor x for which we are contemplating the substitution of y (one hectolitre of grape juice maturing to time $t + 1$). $F_y - F_x$ can now be interpreted as the marginal product of one unit of time, whilst $p(1 + r)^t = mF_x$.⁷ Formula (xvi) can thus be interpreted as the proportional rate of increase of the product with respect to time. This is the Jevons-Wicksell definition of the real rate of interest. Our method of derivation makes it clear, however, that what looks like an attractive generalization and a partial justification of the existence of an interest charge, is, in fact, not a definition but an equilibrium condition. It is, moreover, heavily dependent on highly artificial assumptions. This is emphasized by Wicksell's own attempt at generalization,⁸ where the mere introduction of an additional factor of production makes the growth of product interpretation impossible.

If we now think of x as labor applied directly and y as labor applied indirectly, we have a model similar to Lange's. Using the fact that in equilibrium $p(1 + r)^t = mF_x$,⁹ we may write

$$r = \frac{F_y - F_x}{F_x}.$$

This is the formula quoted by Lange and attributed to J. B. Clark and Wicksell. It is interpreted as the "*ratio of the marginal net productivity of indirect labor to its marginal cost.*"¹ Again it can be seen that this interpretation is dependent upon the artificial assumptions of the model.

In fact it is useless to seek for any *definition* of the rate of interest as the rate of increase of product due to an increase in the use of capital, whether "capital" is thought of as money or as a stock of physical goods which will allow increased waiting. This is because capital has no real product in the same sense as other inputs. If the firm is out of equilibrium, the marginal product of any input is still precisely defined by the production function, but the marginal product of capital is ambiguous until we specify the factor which it is used to purchase. We must not be tempted in our quest for elegance

7. Cf. Wicksell, *op. cit.*, p. 178.

8. *Ibid.*, p. 182.

9. Cf. equation (xv) above.

1. Lange, *op. cit.*, p. 169.

and symmetry to try to force round pegs into square holes. It is always possible to express r as a "real" rate analogous to a marginal product if we use our equilibrium conditions and substitute them into our money definition of the rate of interest, i.e., *profit divided by money capital tied up*. But the resulting equation cannot be regarded as a *definition* of the rate of interest. It will merely be the expression in terms of the marginal products of other factors, which is equal to the rate of interest in equilibrium. It may well be possible to interpret such equilibrium conditions in various ways according to the nature of the productive process, but all imply the original definition of the rate of interest from which they are derived.

IV

It is enlightening now to attempt to fit the work so far into the traditional neoclassical framework. Typically, the notion of a production function is illustrated with equi-product contours. Equi-outlay curves are introduced to show how a scale line can be drawn defining optimum factor proportions at each level of output. This makes it possible to derive a cost curve which, together with a demand curve for the product, is used to identify the equilibrium of the firm.

With our production function there is no difficulty about drawing the equi-product contours. Inputs are measured in physical units and the production function is given in terms of these units. In the two factor case the traditional equi-outlay contour is defined by the equation

$$\begin{array}{l} \text{Total expenditure,} \\ \text{on the average,} \\ \text{per unit of time} \end{array} = \frac{px}{l_x} + \frac{qy}{l_y} = \text{Constant.}$$

The slope of this line is $(p/l_x)/(q/l_y)$. In equilibrium, where equi-outlay and equi-product contours are tangential, we have $(p/l_x)/(q/l_y) = F_x/F_y$. This is the static textbook result.

Armed with our definition of capital it is a simple matter to obtain by an identical procedure results which *look* like those of the Austrian School. Assume that r_o is the market price per unit of time for capital \bar{C} , and that r_o is given and fixed. The constant outlay curve is then $(px/l_x) + (qy/l_y) + r_o\bar{C} = \text{Constant}$. The slope of this line is

$$\frac{\frac{p}{l_x} + r_o\bar{C}_x}{\frac{q}{l_y} + r_o\bar{C}_y},$$

which at the point of tangency with an equi-product contour equals F_x/F_y . Now it follows from the work of Section III that the substitution of our expressions for \bar{C}_x and \bar{C}_y in this last equality reveals that equilibrium requires factor prices to be proportional to corresponding marginal value products *discounted this time at the market rate of profit* r_o .

As indicated, this looks like, but is not, the Wicksellian solution. In Wicksell, the rate of profit is always that which will just exhaust the product.² The aim of the producer is to maximize this profit rate. But the neoclassical entrepreneur takes r_o as given and minimizes cost at the given level of output. In this procedure there is no suggestion that total revenue is exhausted by the capital charge. If it is not, then the residual can be applied to increase r_o . This will immediately affect the slope of the constant outlay curve and hence the equilibrium factor proportion. The neoclassical procedure gives the correct solution only if the aim of the firm is to maximize a residual profit which accrues to an independent "entrepreneur," or if the market rate r_o happens to be the same as the maximum internal rate. Equality of r_o and r_{max} will occur only if both firm and industry are in equilibrium.

From a neoclassical point of view the assumption that it is the profit rate which is to be maximized gives rise to a fundamental difficulty. If r is to be a variable chosen so as to exhaust the product, we can no longer draw a meaningful equi-outlay curve; for this simply becomes identical with the equi-product/equi-revenue curve. The only procedure open to us is that of Section II above. Here we conclude that the best factor proportion is dependent on the maximum internal rate. But this charge varies with output and it varies with the price of the product. It follows that until the price of the product is known, the position of the scale line is indeterminate, and hence that the cost curve itself is dependent on the price of the product.³ Or, more strictly, we have at least two alternative definitions of a cost curve to a firm. The first is that which maximizes the rate of return for each level of output. The second is that which assumes some objective charge for capital which is supposed to be constant for every level of output, and minimizes cost (including the capital charge) for each level of output. These two cost curves are not in general the same. They will both reach a minimum at the same level of output if and only if the objective charge for capital

2. *Op. cit.*, p. 181, equation (1).

3. Cf. Gabor and Pearce, *op. cit.*, p. 256, footnote.

in the second curve is the *maximum maximorum*, i.e., the greatest which it is possible to earn at *any* level of output.

V. SUMMARY

We have achieved what we set out to do; namely, to reveal precisely the differences between the treatment of capital and factor combination in the Austrian and neoclassical theories of production, and to show that the results of these and some other theories are all special cases of the general solution which we developed. We have also derived an exact definition of profit which helps to refute the lingering view according to which the real rate of interest arises because time is in itself productive. What we consider the particular merit of our treatment of the problem is that it leads to a better understanding of business practice. We intend to show, in a later paper, that the approach of businessmen and their accountants to investment problems is fundamentally in line with the theory of profit rate maximization.

THE UNIVERSITY OF NOTTINGHAM

THE AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA

CONTROLLING CONSUMERS IN THE EVENT OF FUTURE WARS

By STEPHEN ENKE

I. Introduction, 558. — II. Some alternative wars and their economic impacts, 558. — III. Alternative suggestions for an over-all expenditure ration, 563. — IV. A pessimistic but realistic program, 569. — V. Concluding comment, 572.

I. INTRODUCTION

The rationing and price control schemes of belligerent governments during World War II stimulated several alternative proposals by economists.¹ However, most of them were ignored by those in authority, at least in the United States, perhaps because they were offered too late. Since 1945, with one notable exception,² there have been almost no carefully considered programs for limiting private consumption and diverting resources to government use in the event of another major war. Also, during recent years, the nature of the wartime problem has become drastically changed, for the United States is now subject to heavy thermonuclear attacks in the event of an "all-out" war; accordingly, one urgent task of government is to plan today the consumer controls needed in anticipation of attack on this country. Also, there are other conceivable wars, somewhat resembling those of the past in their economic impacts, to which formerly suggested rationing schemes may still be applicable.³

II. SOME ALTERNATIVE WARS AND THEIR ECONOMIC IMPACTS

Obviously no one can well describe all the possible wars involving the United States that might occur during the next decade or so.

1. See Section III below.

2. T. Scitovsky, E. Shaw, and L. Tarshis, *Mobilizing Resources for War* (New York: McGraw-Hill, 1951).

3. A few *general* references may be useful for those who have devoted little attention to the subject of war economics. One of the best is *Studies in War Economics*, Institute of Statistics, Oxford University (London: Blackwell, 1947). Also of interest are H. W. Spiegel, *Economics of Total War* (New York: Appleton-Century, 1941); Horst Mendershausen, *The Economics of War* (New York: Prentice-Hall, 1943); E. Stein and J. Backman (ed.), *War Economics* (New York: Farrar & Rinehart, 1942); L. V. Chandler and D. H. Wallace (ed.), *Economic Mobilization and Stabilization* (New York: Holt, 1951); J. K. Galbraith, *A Theory of Price Control* (Harvard University Press, 1952); K. Knorr, *The War Potential of Nations* (Princeton University Press, 1956), and A. C. Pigou, *The Political Economy at War* (New York: Macmillan, 1941; originally published 1921). More specific citations are listed subsequently.

Actually, for this analysis, the only wars that are relevant are those having such a serious impact upon the economy of the United States that the federal government would seek to impose drastic mobilization and rationing controls. And of this set of wars, two subsets must be distinguished: those in which the United States' economy remains physically intact, and those in which its major cities and a substantial fraction of its population are destroyed.

Budget Impacts Assuming No Attacks on the United States

Some approximate idea should be given of what is meant by a war having a substantial impact upon the domestic economy, assuming no attacks on the United States. Today, towards the end of the fifties, the federal government is taking from 16 to 17 per cent of the gross national product, and military expenditures (including those for the Atomic Energy Commission and foreign military assistance) amount to slightly over 10 per cent. Most economists would probably agree that these military expenditures could be *sustained* at twice the current level without recourse to "mobilization" and "rationing" schemes any more specific than those of government spending and taxes.⁴

A more difficult question is how rapidly the federal government could *increase* its expenditures for military purposes without encountering specific shortages that would cause it to control private purchases or pre-empt goods. During the Korean War, although admittedly from a low base of about \$15 billions, Department of Defense expenditures increased threefold within two years, and at a time of high employment. Almost a decade later, the potential GNP is now larger in constant dollars and there is some slack in employment of resources, so that it seems reasonable to suppose that annual increases in military expenditures of from \$5 to \$10 billions each year, for several years, can be financed from taxes.⁵

If these surmises are valid, it means that there is no need to establish new federal agencies to mobilize and ration unless the nation is confronted with an emergency that requires federal expenditures to increase annually for several years by more than \$10 billions to a level exceeding about \$100 billions.

4. The United Kingdom government has taken a quarter of the national GNP in peacetime by traditional fiscal means.

5. This is not to say that the entire amount could or should be raised from existing taxes. It might be necessary to tax articles, activities, and payments that are now untaxed. A wide range of goods are still produced and sold in the United States that might be made subject to a purchase tax; this would also serve to discourage production of consumer goods during a period of national emergency.

Economic Impacts Assuming Atomic Attacks on the United States

It is always possible, of course, that an "all-out" war between the Soviet Union and the United States might occur on very short notice, and not after a year or so of hostilities in other areas of the world. In this case there would probably be no large-scale preliminary mobilization. Economic controls, if there were still an effective government to impose them, would then follow a heavy nuclear interchange between the two powers and would be designed primarily to restore the economy. Economic demands resulting from reconstruction and rehabilitation would be unprecedented. The question then is not whether mobilization and rationing measures are required, but rather whether suitable ones can be prepared before the need for them arises. Such a program must, of course, be operable in a post-attack period.

In the event of an "all-out" war with the Soviet Union in the sixties, it seems probable that the United States will "lose" at least 100 to 150 major towns and cities. The survival of no city having a population of over say 100,000 people can be counted upon. In this connection "lose" probably means that the city can make no contribution to the economy, that few of its buildings and houses are habitable, and that two-thirds of the population are dead, will soon die, or are hopelessly incapacitated. For cities receiving a bomb or missile in the megaton range of yield, the area of almost complete physical destruction may be roughly a five-mile radius. The danger from radioactive fallout will be serious for several hundreds of miles "down wind" from multiple target areas, and with prevailing westerly winds it may be necessary for initial survivors in the Atlantic Coast states to remain under cover for days or weeks following a major attack.

The Soviet Union might attack not only population targets but also some specialized economic facilities, such as petroleum refinery and steel mill centers. The loss of petroleum refineries and adjacent tank farms would have a serious impact on both railroad and truck transportation. Otherwise transportation — except for port facilities — is hard to "target"; railroad track and highway routes, together with rolling stock and motor vehicles, will survive, except within cities. Also attacks might be made on hydroelectric dams and generating stations, of particular importance in the western states.

Although the prospect of losing all major United States cities, and perhaps a third of the national population, is certainly an appalling one, much will survive. Although over half of the industrial capital of the country may be destroyed or unusable, over half the

population and labor force may endure. While a very high proportion of the managerial and technical classes will be lost in city attacks, some will be in rural areas and live. A large fraction of all equipment and labor used in heavy construction and in the extractive industries will survive. The direct impact of nuclear attacks on agriculture, except in areas of heavy fallout, will be slight, and in any event most of the urban demand for foods and fibers will have gone; however, agricultural output will still depend on petroleum products, probably fertilizers, and steel for canning. The people who survive will also be an important depository of social capital, and will not have forgotten how things are done; the ex-urban population of the United States is not an uninformed peasant class.

Postattack Economic Problems

Postattack economic problems are, of course, numerous and onerous. A few of the more important and less familiar ones may be worth mentioning. There is little need to dwell on the obvious need to reconstruct and rehabilitate.

Maintaining Productive Specialization. There is a grave danger, during the panic and disorder following heavy attacks, that people will seek their satisfactions directly, perhaps even by pillaging, rather than indirectly through the economic system. Today the practice of productive specialization is taken for granted. However, in the aftermath of nuclear raids, will these people continue to report for work that satisfied their immediate wants so indirectly?

Securing Government Income. Continuation of some government functions is necessary unless the country is to slip into a state of anarchy. To some extent federal and state governments can commandeer property and conscript unwilling labor. However, if governments and their immediate contractors are to hire labor and buy supplies, they must give something of value in exchange. The problem then is to prearrange matters so that the government has goods or money of value to use as means of payment. It is true that, during World War II, the United States Treasury financed itself as much or more from new money credits as from tax receipts, and to some extent this can and will be done again, but public acceptance of all federal obligations may weaken seriously if it is generally known that the government is collecting few taxes. And the government may find it hard to collect corporate and personal income taxes when private assets are destroyed; many corporate, individual and bank records are lost; and many people are seeking to support themselves directly.

Uncertainties Regarding the Ownership of Wealth. Of less immedi-

ate concern following nuclear attacks, but a problem which over time will become increasingly troublesome, is that there will be great uncertainty regarding the ownership of wealth. Real property valued in hundreds of billions of dollars will have been destroyed, and tens of millions of people will have been killed. There will be assets without ascertainable owners and people without knowledge of their property. Determining successive testators will often be impossible. An enormous number of firms will be bankrupt on paper. Life insurance companies will be confronted with unprecedented claims, dwindling premium income, and assets secured with damaged property.

Unsuitability of Traditional Price Ceilings and Specific Rations

Enough has been written elsewhere regarding traditional price controls and commodity rations to indicate that they are probably inappropriate in the postattack period of "all-out" atomic war.

Traditional controls, with their frozen maximum prices, are predicated on an approximate continuation of past market conditions; if demands and supplies vary markedly in the schedule sense, historic prices are no longer valid. Institution of specific rations for a wide range of commodities and consumer units is an administrative task that, in the postattack confusion, may be beyond the abilities of government officials and local volunteer boards. The enforcement of price ceilings presumes a situation in which most buyers have a long-run interest in preserving the purchasing power of their money, and so to some extent can be counted upon to report violations, but in a bombed-out economy would-be buyers are likely to be desperate to acquire needed goods at once. Much of the public support for traditional controls can be associated with a belief that a war can be carried on concurrently with business more or less as usual, and that the important thing is to prevent large and haphazard income effects upon households. Considerations of this kind will count for little at a time when tens of millions have been killed and more millions of temporary survivors are struggling for the necessities of life.

Last, but of significance, price ceilings and specific rations do almost nothing to resolve the three problems of economic aftermath discussed above.

Nevertheless, the major "Western" powers engaged in World War II, e.g., the United Kingdom, the United States, Australia and Canada, all had maximum price regulations based on some historic period and all rationed commodities to households in more or less specific quantities. Accordingly these are the kinds of controls that the public will probably expect in the event of another war of any

consequence, and which lawyers will always draft without consulting economists. Hence there is a need now to discuss alternative regulating schemes that better satisfy fundamental economic objectives. These goals are: (1) adapting production and consumption according to the ever-changing capabilities and requirements of a war economy that is under enemy attack; (2) allocating each commodity so that it will become re-employed in its new and most important uses; (3) devoting a minimum of resources to the production of goods for civilian use; (4) protecting the very poor from a reduction in real income; and (5) keeping the necessary administrative burden at a minimum.⁶

III. ALTERNATIVE SUGGESTIONS FOR AN OVER-ALL EXPENDITURE RATION

In wartime, when aggregate consumption and hence differences in individual consumption must be reduced, the price system alone cannot be trusted to decide how civilian goods should be distributed among households because its results unduly reflect income inequalities as well as variations in need. However, commodity rationing can overcome income differences only by a wasteful disregard of the peculiar needs of each person and the altered wants of a wartime economy. Accordingly, there have been a number of suggestions for the establishment of some sort of over-all money or point ration to be expended by households on consumer goods.⁷

The essential problem is to arrange matters so that each household, making allowances for its size and composition, has about the same total command over consumer goods. This means that the "money" used by consumers to purchase goods must be rendered distinct from funds that they receive and disburse as entrepreneurs or men of property. And all this must be done in such a way that, in addition to satisfying the five criteria already enumerated, con-

6. The following references provide a fair idea of traditional controls in action: Arthur R. Burns, *et al.*, *Operating under Wartime Marketing Restrictions* (New York: American Management Association, 1943); William A. Nielander, *Wartime Food Rationing in the United States* (Baltimore, 1947); Jules Backman, *Rationing and Price Control in Great Britain* (Washington: Brookings, 1943); Great Britain, Ministry of Fuel and Power, *Committee of Enquiry into Evasions of Petrol Rationing Control* (London, 1948); Canada, Wartime Price and Trade Board, *World Inflation with Special Reference to Price and Supply Controls* (Ottawa, 1944); Australia, Rationing Commission, *Rationing in Australia* (Melbourne, 1944).

7. The leading examples are M. Kalecki's general rationing plan (Bulletin Institute of Statistics, Oxford, Jan. 11, 1941); W. Allen Wallis' expenditure ration and progressive consumption tax (*American Economic Review*, XXXII (Sept. 1942)), and the establishment of white and black bank accounts for consumers by T. Scitovsky, E. Shaw, and L. Tarshis (*op. cit.*).

sumers have an incentive to make the greatest possible contribution to production.

Kalecki's General Rationing Plan

Kalecki suggested in 1941 that the government, instead of setting money and point prices for various goods and establishing specific or point rations for various groups of commodities, should combine these into a single expenditure ration covering all goods bought by consumers at retail stores.⁸ He did not consider administrative difficulties to the extent that Wallis, Scitovsky, and others did in later proposals. He did not concern himself with the possibility that large sums might be diverted to procuring consumer goods not normally acquired at retail shops. As his proposal is less complete than others that will be considered, and as the operation of the scheme was developed less fully by its author, its main interest is historical.

Wallis' Expenditure Ration and Progressive Consumption Tax

In 1942 Wallis proposed a total expenditure ration that would be varied uniformly, depending on occupation, age, household membership, etc. The amount of the ration could be increased by an individual, at a cost, and this would provide some accommodation for persons committed to higher living expenses. This flexibility would come from superimposing, upon the basic total expenditure ration, a rapidly progressing block rate tax on consumer outlays in excess of this sum. The effect, of course, would be to increase markedly the cost to the buyer of purchases made in excess of his expenditure ration. Because it would not be practicable to require individual accounting for all consumption expenditures, it was proposed that all income not saved (as evidenced by investment in United States bonds) should be treated as consumer outlays. Each individual would file a federal income tax in the usual way and the consumption tax would then be calculated by (1) making such adjustments as considered appropriate to arrive at a total net income figure, (2) subtracting purchases of United States bonds and income tax payments, (3) deducting the expenditure ration and any other outlays (e.g., already contracted life insurance premiums) exempted from the tax, and (4) computing the consumption tax on this final balance. This tax would presumably be a temporary war and recuperation tax and it would be in addition to the federal income tax.

From some viewpoints the Wallis plan resembles other proposals. Fixed prices coupled with quantity rations are the equivalent of the

8. *Op. cit.*

value rations for each commodity: and so an over-all expenditure ration can be viewed as the aggregate of many specific price ceilings and specific rations, but with the important superiority that individual preferences can be given a free rein. In many respects, inasmuch as it includes almost *all* expenditures by people as consumers, it encompasses Kalecki's suggestion. Also, it can be regarded as a Treasury scheme to sell bonds, with a quota assigned each person on the basis of taxable income, with penalties levied on the unco-operative.⁹

Although probably an imperfect device for World War III, some scheme of expenditure ration and progressive consumption tax cannot be dismissed out of hand, especially if it were instituted during some Korea-like war and before the onset of World War III.

The advent of Soviet bombers and missiles over the United States, assuming this ration and tax were already in effect, would obviously require modifications in the scheme. Many employing firms that had been making payroll deductions for income tax, bond purchases, and estimated consumption taxes would no longer be in existence. Collection, at probably higher rates, would be very difficult. The Treasury might simply have to instruct all employers to deduct certain large fractions from each paycheck, depending on earned income, and wait until later to ascertain tax liabilities and rebates. Unfortunately these large deductions would, at the time, appear to employees as a considerable tax on pay. Accordingly, there might be a considerable increase in the number of self-employed households, making unrecorded sales of such things as farm produce, and attempting an abnormal degree of uneconomic self-sufficiency.

9. Naturally there are a number of administrative difficulties. Wallis recognized the problem of consumer liquidation of all kinds of assets, and agreed that it would be necessary to guard against liquidation by sale or mortgage to business, banks, or with other money receipts not subject to the consumption tax. Another difficulty is that there will be considerable variation in the debts persons owe at the time such controls as these come into effect. There would probably have to be a moratorium on servicing certain classes of historical debt and provision might be made to exempt from the ration the repayment of certain other commitments. Interest and redemption arising from net obligations contracted subsequent to imposition of the tax would, in most cases, be included in the tax base. Unfortunately it will be impossible to prevent some persons holding large cash hoards at the time that the expenditure ration and consumption tax are introduced; while this superiority endures, consumption goods, new and old, will be in part distributed according to wealth held in the form of cash balances. Also, persons entering the control period with considerable stocks of jewelry, furniture, cars, and housing, etc. will naturally be advantaged during the service life of these goods, whether used by them, exchanged, or sold; however, it is probably in the general interest that these initial stocks of consumer goods be redistributed among users.

It is doubtful whether a productive consumption tax could soon be implemented if none had been in force before the attack. If the government decided, *after* an attack, that it wished to implement some kind of expenditure ration, it would probably be forced into a system not unlike the multiple consumer account scheme to be described. It could make the surviving banks its main operating agents and the limited amount of circulating legal tender its principal instrument of control.¹

An important weakness of the Wallis scheme would be the possibility that retailers, and others receiving currency payments by consumers, will divert a fraction of these for their own use, rather than depositing them in their blocked bank accounts. It is true that some states levying sales taxes have experience in checking sales receipts, but the temptation will be strong and tax enforcement may be weak during this chaotic period. Accordingly, at an early period, the government would probably have to replace circulating legal tender with a new ration money, linked to initial recipients at the time of an authorized withdrawal from a bank, perhaps by perforating the paper notes with the recipient's permit number.

Multiple Consumer Account Proposals of Scitovsky and Others

Many of the difficulties outlined immediately above are resolved — but, of course, at some administrative cost — by a multiple account plan suggested by Scitovsky and others.²

According to this plan a Ration Authority provides each consumer unit with a ration permit, which is a receipt for income taxes, a certificate of income realized during a previous period, and an identification for rationing purposes. The ration permit would be stamped with the consumer's registration number as a precaution against multiple ration allowances for any one consumer unit and as

1. Immediately after attack the federal government could prohibit all cash withdrawals from banks except in accordance with its rules. (All denominations above say \$20 might be recalled as a measure against people who had hoarded currency.) It would be provided, as soon as possible, that all payments by firms to those supplying materiel, labor, or other means of production, be by check. This check would not be cashed, but would have to be deposited by the payee. Consequently, every recipient of income would have to obtain an account, possibly in a new emergency bank. The government would determine the cash withdrawals permitted consumer units each week or month and thus establish an expenditure ration.

2. *Op. cit.* Actually chap. 3 of *Mobilizing Resources for War*, which describes this alternative plan, was written by E. Shaw. In addition, a number of economists at The RAND Corporation, including the present author, contributed parts of the multiple account plan discussed here. Accordingly, a few liberties have been taken with it.

an aid to punitive measures against black market use of ration money. The permit is essentially the community's acknowledgment that the holder has contributed in the recent past to the national product and hence is entitled to buy some share in current national output.

Each consumer unit maintains a white account at some one bank. All consumer purchases are made directly (by writing checks on it) or indirectly (with ration money or small change obtained as a withdrawal). A consumer, assuming he has regular (black) funds to make the purchase, can periodically acquire white balances up to a specified amount. This amount is determined by his bank, which inspects his ration permit and the government table that determines the expenditure ration of the particular consumer unit. (All legal tender except coinage is withdrawn from circulation.) Hence all funds held by a person *as a consumer* are either in the form of white (free) funds, ration money (the equivalent of white funds) or black (blocked) funds.

Blocked (black) consumer balances are obtained in various ways. The most important one is income from productive services such as labor, the loan of money, the lease of land, etc. Bank deposits held at the onset of the control scheme would be transferred to black balances. Other sources of black funds are the sale of securities, the realization of mortgages, inheritances and the like. Black balances are used for payment of financial obligations; that is, for payments not involving a purchase of consumer goods or of resources for the production of consumer goods. To prevent black balances seeping into unauthorized consumption, it might be required that no consumer's black balance can be debited on an interpersonal transfer unless there is a credit to the payee's black account.

Retailers and others who sell to consumers receive ration money, coinage, or checks on white accounts. These must be deposited by each seller, in his firm's *blue* account, held at a single bank. All normal operating expenses of all business firms, including wages, are paid from these blue accounts. Payments to householders for services rendered — as distinct from payments to other firms — are credited to black accounts. Interfirm payments on current account, including payments by retailers to other firms for stock in trade, etc., are paid from blue accounts into blue accounts. (Business firms also maintain *red* accounts for *capital* transactions.)

Households, in short, earn money by contributing to production. These funds are paid into their black account.³ A certain fraction of

3. Actually, in *Mobilizing Resources for War* these are paid into white accounts, and consumers *qua* consumers can only draw on these up to the amount

these can periodically be transferred to a white account which is the source of funds for consumer purchases. The remainder, which may not be so transferred, cannot contribute to the inflationary "gap." Consumer expenditures are paid into the blue accounts of firms selling goods and services to consumers, and these are used to cover operating expenses.⁴

In a war of survival, it is to be expected that annual transfers to consumers' white accounts will be considerably below their annual earnings, as reflected in payments into their black balances. These last will accordingly accumulate fairly rapidly unless withdrawals are made for taxes, government bonds, or other investments. To prevent the latter contributing to inflation and diverting resources from the war effort, there would almost certainly have to be some form of investment control.

This multiple consumer account system is sufficiently novel and restrictive that it is unlikely to be put into force prior to an "all-out" war. However, in the event of any prolonged and serious international tension, it is hoped that consumer units will be registered in advance, ration permits will be issued even if not used, and the various accounts for consumers and firms will be established at least in embryo form. Duplicate ration permit and bank account records should be maintained in physically secure places, on magnetic tape, outside major cities. Ration money should be held in readiness at various points. Public education in the operation of the multiple account plan should be begun. Implementation of the multiple consumer account scheme will be difficult in any event, but it may be impossible without these preattack measures.

Assuming that the plan is established, and is administered successfully, of an unused expenditure ration. However, this really means the establishment of a subsidiary account within the white account. It would seem to be administratively simpler to allow all permitted transfers to white accounts to be used for consumption.

4. Housing and durable consumer goods such as automobiles are a perennial difficulty for all rationing schemes. The government would probably have to ascertain the house or automobile ownership of each consumer unit. Rents from houses owned but not occupied would have to be paid into the owner's black account. Consumer units occupying their own homes would probably receive a smaller expenditure ration. The proceeds of any car sale by a consumer that left it owning one or more cars should probably be paid into the seller's black account. Extra houses and cars owned by a consumer unit might be susceptible to government purchase with black money. Private yachts and airplanes might be handled in the same manner. Most other durable consumer goods are either of modest value or their ownership is not registered in peacetime. In all these latter cases consumer units might be permitted to buy and sell consumer goods for white balances, or of course exchange them. Such transactions would not increase the enforceable claims of consumers as a whole on the current production of the economy.

cessfully, what is its economic value? Does it satisfy the five criteria originally listed (see p. 563)? Does it contribute anything to the three special postattack problems mentioned above (see pp. 561-62)?

Certainly, the specific spending of expenditure rations by consumers being uncontrolled, those resources still devoted to satisfying consumers should adapt more readily to new demand patterns. There being no specific rations, consumers will be more likely to buy what they want, rather than draw their various rations and then try to make exchanges. By limiting white balances and ration money, the government can, in principle at least, reduce to a very low level the diversion of resources to satisfying consumers. As prices should remain low, the poor who cannot work will be protected as most or all of their limited incomes is withdrawable in ration money. Whether the administrative burden, which will fall primarily on the banks, is less than that of traditional controls is hard to say: certainly the public should lose less time waiting when applying for different specific rations.

Multiple consumer accounts, unless supplemented in ways such as described in the next section, may make only a partial contribution to the special postattack problems already mentioned. Unless ration money can command many essentials, and at a price that represents less work for the buyer than if he sought to obtain them directly by his own efforts, the public will tend to abandon specialized production. The government's ability to secure adequate taxes is also dependent upon households using the marketplace and satisfying their wants indirectly by working for others; if this condition is met, and assuming output per capita is not down to the subsistence level, multiple consumer accounts should effectively divert gross national product to government enterprises. Lastly, it is clear that no rationing scheme per se can resolve the third postattack problem of uncertain ownership of wealth, and this is really a subject in itself; however, inasmuch as most money assets of households will be held in black balances, which can only be converted into white money in rationed quantities, sales of assets for black balances will not effect the distribution of consumer goods. Obviously, in view of the widespread confusion, panic, and shortages that can be expected during the postattack period, any ration scheme must be supplemented with a positive program of production and distribution.

IV. A PESSIMISTIC BUT REALISTIC PROGRAM

It is not unrealistic to suppose that the nation may find itself in an "all-out" war, under nuclear attack, without any prior household registration, issuance of ration permits, or establishment of

embryo multiple accounts in local small town banks. On the other hand, because it means government spending and does not inconvenience voters in peacetime, it is quite possible that the federal and state governments may have established large dumps of petroleum products, preserved foods, and other stocks outside major cities. There should also be stockpiling and distribution of "camping" type items.⁵

These stocks, assuming the government prepositions them outside target areas and protects them with force if necessary, could constitute one of its greatest immediate postattack assets. In addition, it can "purchase," with promises to pay, many consumer items in the undamaged areas, and move these to deficit areas for distribution. It is important, except in very small quantities and immediately after enemy attack, that it does not *give* these valuable things away. It should sell them, not for regular legal tender, but for a new war emergency money (WEM) that is already printed and difficult to counterfeit.

The WEM would enter circulation as government payments, and particularly as wages to people working on vital reconstruction projects. Thus work for the government, and perhaps for some of its contractors, is in effect rewarded with a money that can be used at government stores to buy things that most people want urgently. To the extent that a worker does not want these items, but prefers things that can be obtained at regular stores or through private sale for ordinary money, he should find a ready market for his WEM. The buying and selling of WEM for regular legal tender should be encouraged by the government, which could supervise the exchange and publicize the going rate. It does not undermine the scheme if a worker, or someone else holding WEM, uses it to acquire consumer goods through private sale; in fact, the longer WEM is in circulation before it is presented at a government store, the more "free" work and goods the government has had.

Logically, the innovation of WEM by the government is not absolutely necessary; but psychologically it may be a very useful innovation. The government can better popularize its policy of selling prestocked emergency supplies, rather than giving them away, if it sells them for a money that was initially earned on important reconstruction projects. It is true that survivors happening to hold large amounts of legal tender at the time of enemy attack will be advantaged if they can buy WEM for cash, but the price will be high,

5. E.g., tents, sleeping bags, oil stoves, blankets, foul weather gear, axes and shovels, tarpaulins, lumber, nails, common hardware, etc.

and the earner of the WEM will be happy with what seems a high dollar wage. The government is probably anxious, too, to make it appear that the prices of many essentials are not rising very rapidly, and this it can do to some extent by keeping WEM dollar prices in government stores and WEM dollar wages on government-sponsored projects at something like the prewar norm: if there were no WEM, and no specific rationing, these wages and prices would probably rise to very high levels, and the government would appear to be leading the inflation of prices and wages.

Irrespective of whether WEM is introduced, the essential feature of this pessimistic but realistic program is that the government *sells* needed things that it has prestocked, produces after the attack, or "buys" with deferred means of payment. This is a difficult decision for a supposedly representative government to make, especially if the war is continuing and public support is vital, but the government must have some means of obtaining labor and materiel other than conscripting and commandeering. The government must fully exploit, by exchanging them for labor and other materiel, all goods of value that it can control. Most of the taxes that are most financially productive in normal times will now have a very restricted tax base or be uncollectable. Hence, if the government is to remain an effective organization, it would seem to have no alternative, during the first and worst days following major enemy attack, but to make people pay for the goods the government can provide.

During this period the government should be pursuing a hard-hearted but right-minded policy in related economic areas also. Its fundamental aim must be to promote current output, which means that workers must be currently and adequately rewarded and that at least the out-of-pocket expenses of currently produced goods be covered by their prices. As petroleum refineries, steel mills, and other essential industries return to production under government control, the government should try to arrange matters so that its sales receipts are considerably in excess of its current disbursements. In part it can do this by using its newly acquired industrial monopolies. But it can also contribute to its current cash surplus by paying a rent to the owners of surviving capital, used in these industries, that is in some such deferred form as blocked government bonds or black bank accounts. To the extent that the government also uses deferred payments to "buy" existing wealth having a ready market — e.g., extra family cars and houses unoccupied by owners — it can also increase its current surplus. This "profit" — obtained from what in effect is a temporary and involuntary loan — is needed, of course,

to finance "unproductive" government enterprises such as general administration, reconstruction, and perhaps a continuing war effort.

With the passage of time, it should be possible to establish white and black bank accounts in a single bank for each surviving consumer unit. Ration permits would then be issued. The moratorium originally imposed on all bank withdrawals would be eased by the establishment of expenditure rations for all consumer units, as prescribed in the multiple consumer account plan, with some transfers being made from old bank accounts (now a so-called black account) to new white accounts. Blue and red accounts would be established for firms for the purposes already described (see p. 567).

Eventually, assuming WEM had been used during the worst days, it would probably be desirable to recall it, crediting holders' white accounts with the equivalent dollar value as indicated by the free exchange rate. Circulating legal tender in the hands of consumers might also be made white money.⁶ By these successive steps the government, assuming no multiple consumer account scheme had been in effect before enemy attack, might arrive at much the same system; in addition, during the immediate postattack period, by "selling" some of the economy's real capital in return for current productive effort, it may have ensured its own and the nation's survival.

V. CONCLUDING COMMENT

In planning consumer controls for future wars it is well to recognize the political impossibility of introducing in peacetime the sort of scheme that best suits the worst conceivable wartime situation. Hence it is unlikely that a multiple consumer account system could be introduced during a preattack period unless some Soviet aggression, for the time being localized, convinces most people that an "all-out" war is not far in the future. Hence, the best practicable course may be to prepare for the "pessimistic but realistic" program described above, and plan to put some sort of multiple account scheme into operation not long after the start of a larger war.

For wars that can apparently be kept local, the economic strains are not likely to be so great that multiple consumer accounts are necessary; in these cases, assuming frequent and periodic payroll deductions, something like the expenditure ration and progressive consumption tax of Wallis, described here, would seem to be more

6. It is presumed that, as a measure against preattack cash hoarders, all notes exceeding say \$20 were formerly declared valueless unless deposited within a time limit in a bank account previously under moratorium and now blocked for many purposes as a black balance.

than adequate. However, many local wars may not be sufficiently costly to warrant this degree of interference with private economic life.⁷

In concluding, it may be worthwhile briefly to consider the possible consequences of having no scheme ready — other than price ceilings and specific rations perhaps — prior to attack. Under these circumstances, and with reduced tax collections, the government, however much it may conscript and commandeer, will be forced to create new money to carry on its affairs. Without means to block or subtract excess purchasing power, any price ceilings that may have been imposed will disintegrate. Cigarettes and various other commodities may come to constitute the effective currency. Eventually, the accelerating inflation will decimate the investing classes of America, just as former wars have done in Europe. In view of these alternatives there would seem to have seldom been a case where advance planning and limited preparation were so important to the nation.

THE RAND CORPORATION

7. And under no circumstances do traditional price ceilings and specific rations seem to solve any real economic problem, although their political *raison d'être* is obvious enough.

COMMERCIAL POLICY AND ECONOMIC NATIONALISM

By FRANK H. GOLAY

- I. Theory of economic policy and processes of economic growth, 574. —
II. Factors in appraising the developmental content of economic policy, 577. —
III. Commercial policy in postwar Southeast Asia, 579. — IV. Conclusion, 586.

For the past three-quarters of a century, a succession of perceptive observers of colonialism — ranging from Rudyard Kipling and George Orwell through J. S. Furnivall and M. L. Darling to economists of the stature of S. H. Frankel and J. H. Boeke — have been calling attention to the “sea change” which Western economic institutions undergo when they are transplanted in traditional tropical cultures. This paper examines the transformation of commercial policy under the impact of economic nationalism in Southeast Asia.

I

International trade theory and more specifically the theory of commercial policy has occupied an important position in speculation regarding the processes of economic growth. The case for free trade, i.e., “neutral” commercial policy, was a case for international specialization, an “optimum” allocation of resources and the maximization of the real income of the world. Opposing the case for free trade there evolved a theory of protection, i.e., “positive” commercial policy, which assigned national objectives to commercial policy including that of accelerating national economic growth. It was contended that “positive” commercial policy would influence growth through the following independent relationships. First, protection could be used to raise the prices of imports and import-competing products produced domestically, thereby establishing an incentive for private developmental expenditures in the protected industries. Crucial to the theory of protection was the argument that the structure of international specialization was in large part historical accident and that recourse to protection would enable late-comers to avoid specialization unfortunate for the welfare of the nation. Second, taxation of imports would produce government revenues which would be available to finance developmental expenditures, i.e., social investment. Finally, commercial policy could be used to affect the distribution of the gains from trade in favor of the country resorting to protection.¹

1. Maximum protection will minimize revenues from import taxation and ultimately will minimize the gains from trade.

The traditional theory of protection relied upon import (and export) taxation for implementation. The claim of the state to the windfall arising out of import restriction was unquestioned. Moreover, any monopoly returns to domestic producers of import-competing commodities arising out of import restriction would tend to disappear over time as output increased. Finally, the improvement in the terms of trade which might result from protection would be distributed widely through society. In the relatively stable environment of the international gold standard, the case for taxation of foreign trade to implement a "positive" commercial policy was persuasive.²

Speculation with respect to commercial policy has kept apace with the far-reaching changes in economics of the last quarter century and recent years have seen the ascendancy of a nationalistic commercial policy of controlled external disequilibrium. The objective of such a commercial policy in underdeveloped countries is to maintain a tolerable equilibrium in external payments while pursuing an economic development policy implemented by a high level of investment. It is recognized that increments of investment even where financed by additional domestic private savings or taxation will contribute, in the short run, to external disequilibrium if there is an import component in the additional investment.³ Moreover, there is an important role to be played by deficit finance and private credit expansion in accelerating economic growth, and it is widely recognized that the controlled disequilibrium of developing countries should properly arise in part from a high level of investment not matched by comparable levels of intended domestic savings.⁴

Underdeveloped countries are handicapped in their efforts to accelerate economic growth by low levels of savings, inelastic supply

2. Symmetry of analysis of traditional commercial policy requires consideration of export taxation as well as import and export subsidization. However, theoretical analysis as well as implementation of commercial policy has been dominated by protection, i.e., import restriction.

3. Underdeveloped countries have found that additional investment financed wholly or in part by foreign loans or economic aid may contribute to external disequilibrium where the increment of investment involves domestic expenditure as well as foreign exchange expenditure. The increased incomes resulting from the increment of domestic expenditure together with the import content of the investment may generate a demand for imports of goods and services in excess of the foreign exchange proceeds of the foreign loan or economic aid.

4. A recent comprehensive statement of the theory of developmental commercial policy is to be found in G. Myrdal, *An International Economy* (New York: Harper Bros., 1956), pp. 267-88. The footnote references of this work provide a useful bibliography of the extensive literature treating the relationships between commercial policy and economic growth.

functions, relatively large marginal propensities to import, and limited amounts of foreign exchange reserves. Under conditions of unlimited foreign exchange reserves, the investment multiplier in the typical underdeveloped country might not be large inasmuch as the small propensity to save would be counterbalanced by a relatively large propensity to make current account payments. However, in the absence of foreign exchange reserves or foreign financing adequate to sustain a large deficit in the current account, the investment multiplier tends to be large as governments are forced to limit foreign payments.

The underdeveloped country resorting to exchange and import controls may be faced by the need to intensify steadily such controls. Inelastic supply functions combined with a large investment multiplier tend to (a) generate price increases which impair the competitive position of exports and (b) lead to increased domestic consumption of output previously exported.⁵ The resultant decline in foreign exchange proceeds will require intensification of controls to maintain a tolerable external equilibrium. Moreover, controls over trade and payments have proved to be an obstacle to an inflow of foreign investment which would tend to materialize in the initial stages of development.

In the face of such formidable obstacles, countries attempting to accelerate economic growth impose stringent quantitative controls over imports (and other payments) to maintain a precarious but tolerable equilibrium.⁶ Generally, the rationing of foreign exchange is on the basis of developmental priorities and an increasing proportion of foreign exchange tends to be reserved for imports of capital goods. Such a theory of developmental commercial policy has considerable relevance and many underdeveloped countries can point to their balance-of-payments difficulties as evidence of effective attempts to accelerate economic growth.⁷

However, this neat theory is also the refuge for laggards in the

5. Such changes in prices and costs are not essential to the theory of developmental commercial policy. External imbalance will normally result from the tendency for Southeast Asians to spend a substantial part of increments of income on imports of goods and services.

6. In a formal sense it is incorrect to include exchange controls in commercial policy, since exchange rationing applies to capital transfers as well as to trade and services. However, it is realistic to consider exchange controls in underdeveloped countries as an aspect of commercial policy in view of the tendency for both exchange restrictions at the margin as well as the substantial part of exchange restriction to be applied to current account items.

7. Cf. Myrdal, *op. cit.*, p. 270. "Under these circumstances it should, perhaps first be said that there must be something wrong with an underdeveloped country that does not have foreign exchange difficulties."

race to accelerate economic growth. While external disequilibrium may be a necessary concomitant of economic development, the existence of balance-of-payments difficulties is not necessarily evidence of effective efforts to accelerate economic growth.⁸

II

Assessment of the developmental content of an economic policy must invariably involve subjective judgments. Such an appraisal should reflect, among other things, understanding of (a) the structure of values of the society, (b) the structure of political power and the process by which public policy decisions are made, (c) the resiliency of social institutions incompatible with accelerated economic growth and (d) motivations of personalities who are in a position to influence public policy decisions.

In addition, the following types of statistical evidence would be relevant to such an appraisal. First would be the level of investment and savings when compared with (a) the historical requirements of development, (b) the performance of underdeveloped countries effectively accelerating economic growth, and (c) the potentialities of the society for capital formation in terms of the existing level of income and the concentration of income. For example, investigation of capital requirements of rapid economic progress suggests that net investment of 5 per cent of income may be required to maintain existing levels of per capita income where population is increasing at 1 to 2 per cent per annum.⁹ Second, it is generally recognized that the state has a responsibility to contribute to economic growth by public saving and social investment as well as by creating a framework of policy favorable to growth. On the other hand, such governments are confronted by imperative and expanding demands for public services. Where the proportion of income captured by taxation is not in excess of 10 per cent, the state is probably not in a position to

8. This brief resumé of the theory of developmental commercial policy is not intended to suggest that capital formation is the critical ingredient of economic growth. It is not unlikely that significant changes in values and social institutions will have to precede effective efforts to accelerate economic growth in Southeast Asia. However, the theory of developmental commercial policy outlined here indicates the appropriate policy response for the country attempting to accelerate economic growth by high levels of developmental expenditure and has not been, so far as I know, related to the problem of cultural change.

9. For a comprehensive bibliography of the literature dealing with the capital requirements of economic progress, see: S. A. Abbas, *Capital Requirements for the Development of South and Southeast Asia* (Groningen, Netherlands: J. B. Wolters, 1956), pp. 145-51.

contribute effectively to net domestic investment.¹ Third, the existence of stringent controls over foreign payments in an environment of stable or declining internal prices is hardly compatible with the theory of developmental commercial policy in which the incomes generated by a large investment multiplier in the absence of adequate foreign exchange reserves would be used to purchase domestically produced goods and services. Implicit in such a concept is the belief that the over-all supply function in the early stages of development will tend to be relatively inelastic.

The Philippines represents a prototype of postwar Southeast Asian commercial policy of controlled disequilibrium not arising primarily out of effective efforts to accelerate economic growth. In the four years following 1951, national income totaled ₱29.3 billion. During this period gross fixed capital formation totaled ₱2.2 billion while depreciation was estimated at ₱1.7 billion. When net fixed capital formation is combined with estimated increases in stocks during 1952-55 of ₱.5 billion, it is doubtful whether net domestic capital formation exceeded 5 per cent of national income. Philippine capacity to raise revenues by taxation has also been undistinguished. Tax revenues, comprising between 85 and 90 per cent of government revenues, totaled ₱2.7 billion, or approximately 10 per cent of national income during the four fiscal years ending June 30, 1956. Finally, Philippine prices have remained relatively stable since 1949. For example, the cost of living index (1953 = 100) averaged 99 in 1949 as compared with 98 for 1955.²

The stability of Philippine prices when combined with the other items of evidence, strongly suggests that the precarious balance in Philippine payments which has been maintained since 1949 by stringent exchange and import controls, does not arise out of effective attempts to accelerate economic growth. The need for such controls arises out of an exchange parity which overvalues the peso. The official exchange parity of two pesos per dollar is unchanged from the prewar parity.³ The Philippines emerged from World War II with a structure of costs and prices at least five times as high as before

1. Implicit in this conclusion is the assumption that potential price inflation effectively circumscribes developmental expenditures which might be provided by budgetary deficits or nongovernmental bank credit expansion.

2. An attempt to assess statistically the developmental content of commercial policy lends an aura of preciseness which is unwarranted. The effectiveness of development efforts is a matter of degree, and is not absolute; statistical series can only support, not confirm, the essentially intuitive judgment of policy.

3. The Philippines is the only country that suffered extensive capital consumption and destruction in World War II which has been able to defend its prewar parity.

the war while prices in the United States, which accounted for three-quarters of Philippine foreign trade, were approximately doubled. While the decline in prices in the early postwar years significantly reduced the overvaluation of the peso, the present value of the peso exchanged for dollars at the official exchange rate is greater than the purchasing power of the peso internally.

The selection of the Philippines as the prototype of postwar Southeast Asian commercial policy was due primarily to the availability of statistics. While there are important differences from country to country, with the exception of Malaya,⁴ all of the states of Southeast Asia are maintaining external balance by stringent import and exchange controls. Moreover, the pressure on the balance of payments does not, in general, arise from a high level of developmental expenditure and the effectiveness of public policy in accelerating economic growth.

III

It is the thesis of this paper that commercial policy in postwar Southeast Asia is intelligible not in terms of efforts to accelerate economic growth, but in terms of the economic content of nationalism. The economic counterpart of political nationalism in newly-sovereign Southeast Asia is best understood as a determination to "de-alienize" the economies inherited from the period of colonialism. Commercial policy has become a powerful tool for increasing both absolutely and relatively the ownership or control by "nationals," i.e., ethnic Southeast Asians, of the nonagricultural assets of their economies. It should be emphasized that the newly sovereign governments of Southeast Asia resorted to exchange and import restrictions in response to an intractable imbalance in external payments. The commercial policy of "de-alienization" did not appear fully developed in Southeast Asia but evolved in a process of experimentation and adaptation of commercial policy expedients to the long-run objective of economic nationalism.⁵

The colonial period was distinguished by the imposition of *laissez faire* liberalism on the tradition-dominated cultures of Southeast Asia. Because of their cultural heritage Southeast Asians were poorly equipped to engage in the highly individualistic "jungle capitalism"

4. Including Singapore.

5. The choice of the term "de-alienization" to describe such a commercial policy was made to avoid confusion which might attend the use of "nationalization" as an alternative. The latter term has acquired the connotation of public ownership which has not been initially a dominant objective of economic nationalism in Southeast Asia.

which resulted. They were shunted to the fringe of the market place to witness the "alienization" of their economies — where the aliens included unassimilated Chinese, Arab and Indian minorities as well as the dominant Westerners. These highly rational minorities acquired control of the bulk of the nonagricultural assets largely because of business acumen and a higher propensity to save, qualities richly rewarded in a liberal economic system, but qualities alien to traditional Southeast Asia.⁶ Colonialism proved indefensible in World War II and the period since 1945 has witnessed the triumph of nationalism throughout the area. The political exclusiveness of nationalism is matched by a determination to "de-alienize" their economies.

Restriction of current account payments creates a windfall where the limited imports (and services) are rationed by the price mechanism. A fundamental commercial policy problem is the distribution of the windfall arising out of foreign exchange rationing. The governments of Southeast Asia are confronted by strong centrifugal forces inherent in nationalism in an environment of ethnic heterogeneity. The capacity of these governments to acquire command over resources to meet imperative demands for both welfare and accelerated growth are limited. Under these circumstances, there is a strong incentive for the government to appropriate the windfall arising out of exchange rationing.

First, real resources are diverted to the government to the extent that foreign exchange is utilized for government expenditures. Because of their monopoly of foreign exchange, governments are able to acquire their foreign exchange requirements in exchange for local currency at rates which substantially overvalue the local currency. Second, governments derive revenue from premium sales of foreign exchange allocated for those uses where exchange controls have been applied most intensely and, therefore, have produced the widest divergence between the internal and external value of the local currency. Multiple exchange rate schemes are presently being implemented by most Southeast Asian countries. Most premium rate schemes have the additional feature that they tax nontrade payments which, by definition, escape the tariff schedules. Finally, an overvalued currency facilitates the collection of normal tariff duties, and reduces the opposition to the imposition of higher duties.

While Southeast Asian governments have tended to be more effective in capturing the windfall arising out of import restriction, it

6. Where rationality is assessed in terms of values and attitudes prevailing in the West.

is still true that quantitative controls establish the effective, or marginal restraint, on foreign payments. That is to say, foreign exchange licenses are highly sought after because the tariffs and exchange premia which have to be paid do not exhaust the windfall established by the stringent rationing of foreign exchange. Holders of import licenses are able to sell imports at prices the market will bear and the tendency for imports to expand less rapidly than money incomes in Southeast Asia has tended to sustain the heavy windfall to be realized from foreign exchange licenses. Applications for the available foreign exchange tend to be in excess of allocations and the economic incentive generated by windfalls arising out of import restrictions has motivated a considerable part of the corruption which has plagued the governments of postwar Southeast Asia.⁷

The remainder of this paper is devoted to examination of commercial policy in which a substantial part of the windfall arising out of foreign exchange rationing and currency overvaluation is not appropriated by the government directly, but is used to accomplish "de-alienization."

The initial objective of "de-alienization" has been, naturally enough, the import trade. The windfall created by import restriction has ensured the economic viability of inexperienced indigenous entrepreneurs while the administrative discretion inherent in import and exchange controls has facilitated discrimination against alien firms.⁸

7. The failure of Southeast Asian governments to capture the windfall inherent in import restriction is attributable, in part, to the sensitivity of Southeast Asian nationalism. Exchange premia are recognized as a measure of effective devaluation and, it is believed, would be interpreted by the outside world as evidence of administrative weakness or policy mismanagement.

8. Belief that the producer/exporter in the trading partner will be able to appropriate a major part of the windfall arising out of import restriction persists in economic analysis of exchange and import controls. Cf. Joan Robinson, "Beggars-My-Neighbors Remedies for Unemployment," reprinted in *Readings in the Theory of International Trade* (American Economic Association, 1949), p. 403 and J. E. Meade, *The Balance of Payments* (London: Oxford University Press, 1954), pp. 276-89. Such an alternative is of little practical significance. Exchange control authorities do not link exchange allocations to a specific producer/exporter in a trading partner. The importer is normally allocated foreign exchange to import specified commodities and he is free to choose among alternative sources of supply. At worst, the importer might be restricted to expenditure of an exchange allocation in a particular country.

The above conclusion is subject to two qualifications. First, when foreign exchange is allocated to an alien firm or to the branch or subsidiary of a foreign owned firm it might be argued that the producer/exporter in the trading partner has appropriated the import windfall. However, economic nationalism can be depended upon to police this type of arrangement and to prevent the direct expropriation of the windfall through higher prices paid for imports. Second, where irregular currency overvaluation is maintained by exchange and import

The popularity of such a policy is manifested in the eagerness of political parties and politicians to legislate an increasing portion of the available foreign exchange to indigenous importers.

A co-ordinate objective of "de-alienization" has been the retail trade which, in Southeast Asia, has been dominated by the Chinese and Indian minorities. Here the progress of "de-alienization" has been slower because of the technical problem of diverting a portion of the import windfall to the indigenous retailer. One solution to the latter problem has been the creation of statutory import monopolies which are allocated foreign exchange at the favorable official rate and which in turn wholesale imported goods to indigenous entrepreneurs.⁹ By passing on to the indigenous retailer a portion of the import windfall, the retailer is enabled to compete successfully with alien retail firms which must acquire imports at prices reflecting the severity of import restriction.

Finally, commercial policy contributes to "de-alienization" through its impact on the export sector. The relatively industrialized plantation, mining and forestry industries producing for export prior to World War II were dominated by Western firms. Following World War II, industrial production of tropical exports, handicapped by political insecurity as well as currency overvaluation, has tended to lag behind prewar levels whereas world production and trade have expanded substantially over prewar levels.¹ The postwar period has witnessed a substantial "de-alienization" of industrial export production — a process characterized by (a) failure to rehabilitate war-damaged Western-owned industries, (b) capital consumption forced by the burden of currency overvaluation and stringent controls over capital export, and (c) "distress" sales of Western enterprises to indigenous entrepreneurs.

Similarly, the assembly and shipment of peasant exports, e.g., rice in Vietnam, Burma and Thailand, coconut products in the Philippines and Indonesia, and rubber of small holders in Malaya

controls, the importer allocated exchange to import commodities not severely restricted has an incentive to overvalue imports in order to accumulate liquid foreign assets. Such liquid foreign assets might be accumulated in anticipation of devaluation or in order to realize a premium from transactions in the "free" foreign exchange market. Even where import overvaluation requires collusion of the producer/exporter it is unlikely that the importer would not be able to appropriate the major part of the windfall arising out of import restriction.

9. For example, the Price Stabilization Corporation (PRISCO) and its successor the National Marketing Corporation (NAMARCO) have implemented such a policy in the postwar Philippines.

1. Malaya, which has not rationed foreign exchange, i.e., maintained an overvalued currency, is an exception.

and Indonesia, have traditionally been dominated by Chinese and other Asian minorities. In the case of peasant produce, "de-alienization" has often been accomplished by statutory monopolies staffed by nationals, which procure and market both domestically and in international trade. The origins of statutory monopolies in peasant produce are complex and have varied widely from country to country.² However, in Southeast Asia they have been closely related to commercial policy. The effectiveness of exchange controls depends upon the ability of the exchange control authority to capture foreign exchange proceeds. On the other hand, under conditions of currency overvaluation and external disequilibrium, individuals who come into possession of foreign exchange are motivated to evade the controls. The usefulness of statutory export monopolies in the implementation of exchange control is obvious.³

What further consequences are attributable to a commercial policy of "de-alienization"? I would call attention first to two adverse effects and then speculate briefly regarding two other effects which contribute to a rationale for such a policy.

Economists generally deplore the nature of colonial economic specialization. It is usually argued that excessive dependence on exports of a few primary products adversely affects economic progress both because of a tendency for the terms of trade to deteriorate secularly and because of cyclical instability of prices received for primary products. However, it seems evident that currency overvaluation resulting from commercial policy of "de-alienization" will not contribute to diversification of export production. Such a policy narrows the opportunities for specialization to a smaller spectrum of absolute advantage. The countries of Southeast Asia are today more dependent upon fewer exports than ever before.⁴ In the absence of a change in commercial policy, e.g., export subsidization, premium buying rates for exchange, etc., there is little likelihood that the increasing concentration of exports will be reversed. Export produc-

2. Cf. C. Lebuscher, *Bulk Buying from the Colonies* (London: Oxford University Press, 1956).

3. As is well known, statutory monopolies for the procurement and marketing of peasant produce have acquired important revenue functions. The isolation of peasant producers from "superior" consumption patterns and the strength of tradition in the rural countryside results in inelastic supply conditions of small holders' production. Under such circumstances, statutory monopolies have been able for prolonged periods to maintain procurement prices substantially below export prices without reducing output substantially. Taxation of exports of rice has provided a major part of revenues available to the government of Burma following independence.

4. See table next page.

tion and trade in Southeast Asia are being "de-alienized" by a commercial policy which reduces the volume of trade and increases the product concentration.

Traditional commercial policy achieves protection, i.e., raises the prices of imports by taxation whereas effective exchange and import controls establish a margin of protection by reducing the volume of imports which are marketed at higher prices. There are significant qualitative advantages to protection implemented by tariffs. To begin with, exchange and import controls tend to fluctuate widely in intensity being affected by both economic and political factors. The effectiveness of protection in inducing investment in domestic production is related to its greater certainty and steadiness. Moreover, protection arising out of exchange controls seems to preclude substantial foreign investment. To the extent that exchange controls require conversion of foreign currencies at the official (overvalued) exchange rate to meet local currency costs of investment, such investment will be discouraged. Infinitely more important, the existence of discretionary exchange controls are a formidable obstacle to private foreign investment when implemented in an environment of nationalism. Investment induced by protection afforded by exchange and import controls will tend to be indigenous investment. This is quite consistent with economic nationalism and the objective of "de-alienization," but it may not contribute to maximum economic growth.

Commercial policy implemented by import taxation tends to influence the terms of trade, in the first instance, by reducing the

4.

COMMODITY DISTRIBUTION OF EXPORTS OF SOUTHEAST ASIA¹
(Per cent)

Indonesia	Average		Philippines	Average	
	1937	1953/55		1937	1953/56
Rubber	30	36	Coconut products	28	39.5
Petroleum	17	24	Sugar	38	25.0
Tin	8	8	Abaca	14	8.0
Coconut products	7	7	Timber and lumber	—	9.5
	—	—		—	—
	62	75		80	82.0
Thailand	Average		Burma	Average	
	1937	1953/56		1937	1953/56
Rice	46	53.0	Rice	42	76.8
Rubber	14	18.5	Teak	7	2.0
Tin	23	6.3	Cotton	2	4.3
	—	—	Petroleum	25	—
	83	77.8		—	—
				76	83.1

1. International Monetary Fund, *International Financial Statistics*, X (Sept. 1957), 32-36.

demand for imports, thereby reducing the price received by the foreign producer. After the imposition of tariffs the domestic consumer pays more for imports while the foreign producer tends to receive less while the difference between these prices represents the revenue diverted to the tariff-imposing government. Similarly, in the case of an overvalued currency maintained by exchange and import controls, effective demand for imports is limited by exchange rationing thereby tending to reduce the price received by the foreign producer. Moreover, an overvalued currency impairs the competitive position of export production, tending to reduce the supply of exports and thereby resulting in higher prices for exports. There is strong evidence of a marked improvement in the terms of trade for tropical produce in the postwar as compared with the prewar period.⁵ The high degree of uniformity of commercial policy in the tropical, less developed world has contributed to improvement in the terms of trade of this area. However, these countries have not shared in the postwar expansion in world trade and, given their ambitions for accelerated economic growth, no responsible observer can be complacent over the level of imports obtained by these countries. The improved terms of trade for tropical produce may have been procured at the cost of reduced foreign exchange earnings in real terms.

Finally, it is generally acknowledged that economic growth may be highly dependent upon cultural and social change as well as upon capital accumulation. The highly nationalistic commercial policy of Southeast Asia may be productive of social change. For example, "de-alienization" tends to develop indigenous entrepreneurship. While initially the fortunate holders of import licenses act as "dummies" and peddle their licenses to alien businessmen, their experiences with commercial transactions tend to broaden their horizons while the windfall which they realize from import restriction provides them with capital and in increasing numbers they turn to commerce and other economic activities.

Moreover, the economic pressure on the Chinese minority has led to a discernible tendency on their part to seek and train indigenous

5.

INDEXES OF UNIT VALUES IN NATIONAL CURRENCIES OF IMPORTS AND EXPORTS

	Philippines		Thailand		Burma		Fed. of Malaya	
	Imports	Exports	Imports	Exports	Imports	Exports	Imports	Exports
Prewar	51 ¹	34 ²	6 ²	9 ²	41 ²	10 ²	30 ¹	29 ¹
Postwar	132 ³	118 ³	127 ⁴	124 ⁴	130 ³	111 ³	130 ³	146 ³

Source: United Nations, *Yearbook of International Trade Statistics*, 1955.

1. 1937.

2. Unweighted average 1937-39.

3. Unweighted average 1952-55.

4. Unweighted average 1947-49.

business associates and increasingly to cast their lot with the nation state of their residence. The process of "freezing" alien minorities out of commercial activities combined with exchange restraints on the export of capital has led to a shift in the activities of the Chinese and other minorities toward production of light consumer goods for the domestic market, frequently in association with native entrepreneurs.

Social change is also accelerated by the concentration of activities associated with commercial policy in urban centers. The potential Southeast Asian entrepreneur who responds to the economic opportunity presented by "de-alienization" gravitates to the cities where he tends to shed traditional values and loyalties which are widely recognized as obstacles to rapid economic growth. The economist observing postwar Southeast Asia cannot but be aware of the significant element of urbanized Southeast Asia with consumption patterns and economic functions that are usually associated with the middle class in Western societies.

IV

The commercial policy of a country at any moment in time is a composite structure reflecting diverse public policy objectives. Policy goals which influence commercial policy include, among others, economic development, industrialization, economic nationalism, fiscal expediency, political autonomy, etc. Examination of commercial policy in Southeast Asia indicates that economic nationalism as manifested in a determination to "de-alienize" their economies exerts a predominant influence upon current commercial policy in this area.

Commercial policy of "de-alienization" described herein is subject to stresses which will tend to produce rapid institutional change. For example, currency overvaluation will tend to break down as (a) exporters and others coming into possession of foreign exchange devise techniques for evading exchange controls, (b) pressure for consumption through the budget motivates governments to exhaust import values as a tax base and (c) the drawbacks of existing commercial policy increasingly outweigh the benefits and the need to develop alternative techniques for "de-alienization" is recognized.

The literature dealing with developmental commercial policy tends to dismiss those countries having the symptoms of developmental commercial policy in the absence of effective efforts to accelerate economic growth as "satisfied with the existing relative stagna-

tion.”⁶ Such a conclusion represents an evaluation of Southeast Asian policy in terms of Western oriented values. When appraised in terms of the values and aspirations of Southeast Asian nationalism, commercial policy in that part of the world becomes both rational and dynamic.

The sovereign states of Southeast Asia are determined to reintegrate their societies on the basis of an exclusive nationalism. Economic nationalism up to the present has not been so much concerned with the size of the pie, i.e., economic development, as in increasing the share accruing to Southeast Asians. Initially the windfall arising out of import restriction tends to motivate more graft and corruption than legitimate economic activity, but there is evidence that it may accelerate social change prerequisite to rapid economic growth. In any case, “de-alienization” is a popular policy in Southeast Asia, and understanding of problems of economic growth in that part of the world will improve with recognition by economists of the powerful economic forces inherent in nationalism. A basic public policy problem is the deterioration of the relative position of Southeast Asia in world trade. Stagnation of export production and trade does not have to be a cost of “de-alienization.”

CORNELL UNIVERSITY

6. Myrdal, *op. cit.*, p. 269. Myrdal's dismissal of a substantial part of the underdeveloped world is surprising in view of his avowal (p. 223): “I propose, in discussing the commercial policy problems of the underdeveloped countries, to tackle the subject, deliberately, from the viewpoint of their own interests.”

UNITY IN THE THOUGHT OF ALFRED MARSHALL*

By G. PURSELL

Introduction, 588. — I. The theory of value, 589. — II. Money, interest and fluctuations, 590. — III. Aspects of economic development, 594. — IV. Conclusion, 599.

Marshall came to economics in 1867; within three years he had worked out the theoretical backbone of his system.¹ By the time of the publication of Jevon's *Theory of Political Economy* in 1872 (he wrote to J. B. Clark, almost forty years later) "I had practically completed the whole of the substance of my Mathematical Appendix. . . . Substantially my theory of capital as it exists today is completely outlined in Notes V and XIII–XIV; and my general theory of distribution (except in so far as relates to the element of time) is in like manner contained in Note XXI; to which the preceding notes and especially XIV–XX lead up. . . . Between 1870 and 1874 I developed the details of my theoretical position; and I am not conscious of any perceptible change since the time when Böhm Bawerk and Wieser were still lads at school or College. . . . My whole life has been and will be given to presenting in realistic form as much as I can of my Note XXI. If I live to complete my scheme fairly well, people will, I think, realise that it has unity and individuality."²

No one today would question the "individuality" of Marshall's contribution to economics. But its unity? Although he heatedly denied the suggestion that he tried to compromise, to reconcile divergent theories — "Truth is the only thing worth having, not peace!"³ — our own age is skeptical. There is the "real cost" question, the reverence for Ricardo and the English tradition, the small amount of attention paid to the Austrians and French (whose position was really much closer to his own), the mighty treatise resembling Smith's in so many ways. All this is clear by now; but what has possibly not been discussed so much is the unity of his work, *given* his "real" position on fundamentals. It is suggested in this paper that Marshall's treatment of certain topics was restricted by his desire to maintain this unity, and that discoveries which were important were not fol-

* I am indebted to Professor T. W. Swan and Dr. I. F. Pearce for comments.

1. T. W. Hutchison, *Review of Economic Doctrines, 1870–1929*, p. 64.

2. *Memorials of Alfred Marshall*, ed. A. C. Pigou (London, 1925), p. 417.

3. *Ibid.*, p. 418.

lowed up, as they could have been if he had not so restricted himself.⁴ First there is a description of what is understood to be his central position, his theory of value; then this is discussed in relation to his theory of credit, interest and fluctuations; and finally in relation to various aspects of his ideas on economic development.

I. THE THEORY OF VALUE

It is apparent from the quotation above that Marshall's general theory of value was arrived at very early. In essentials the system was the marginal utility system of general equilibrium, not different from that of Walras, or those of Jevons and the Austrians. The starting point was the "wants" or demands of individuals; these could be satisfied by exchange and production in competitive markets. "Costs" were the "imputed" payments to "services" rendered by "factors of production." It was assumed that all the economic agents acted rationally, in the sense that they so allocated resources between different wants as to maximize satisfaction. The system could be expressed by simultaneous equations, or, more vividly, described by means of the "stationary state." This was the theoretical "long run" where everything is adjusted from the past, and where "the requirements of a future age can be anticipated an indefinite time beforehand."⁵ Of course Marshall did not agree that costs were subsidiary; they had an independent place of their own, like a billiard ball balancing against two others in a bowl. However, it is apparent now that the logic of his position was otherwise; as Schumpeter puts it, "the position of the balls is to be accounted for by a single principle, gravitation in the case of mechanics, utility in the case of economics."⁶

In what sense was this position the center of Marshall's thought? After all, he repeatedly spoke of theory as "an engine for the *discovery* of concrete truth,"⁷ and emphasized the special nature of concrete problems. He distrusted "long chains of deductive reasoning," and the fine adjusting of existing tools, when each real situation would largely require the construction of its own. While this is true of his statements of methodology, in practice the warning to others and to himself is often overlooked; doubtless due partly to "his strong ethical intent and desire to hand on some practical message in the accepted

4. I disagree with L. E. Fouraker ("The Cambridge Didactic Style," *Journal of Political Economy*, LXVI (Feb. 1958)) that the difficulties to be found in the *Principles* are always, or even mainly, a matter of presentation or "style."

5. Marshall, *Principles*, 8th ed. (Reprinted 1952), p. 305.

6. J. A. Schumpeter, *History of Economic Analysis*, p. 922.

7. *Memorials*, p. 159. My italics.

time-honoured terminology,"⁸ but also, it is suggested, because of his unwillingness or inability to use other engines of analysis than his own.

II. MONEY, INTEREST, AND FLUCTUATIONS

The explanation of interest raises difficulties even in the pure theory of the stationary state. In Marshall interest is a payment for the use of loanable funds — liquid funds — as distinct from concrete capital factors. The supply of loanable funds is clearly though perhaps not very adequately discussed. Saving must be rewarded in a stationary state because of man's "irrational" preference for present pleasures; and, in the real world, there is the additional factor of uncertainty (though this is complicated by the simultaneous entry into the analysis of "social and religious sanction," "family affection," etc.).⁹ But on the demand side the position is hazy. Marshall points out the dilemma,¹ but does not anywhere explicitly resolve it. The demand for loanable funds is derived from the demand for (gross) investment. But this in turn depends on the prospective yields of additional capital factors, *discounted* to the present. Discounted by what? It cannot be the market rate of interest, because that is to be determined. The answer must be the individual time-rates of discount of those to whom the capital good yields accrue.²

This, then, is the "engine" with which Marshall proceeded to examine concrete cases. Complementing it is his theory of money, also closely integrated into the general theory of normal value. The value of money, according to this, the "Cambridge" equation, is determined by the demand and supply for it.

"In every state of society there is some fraction of their income which people find it worth while to keep in the form of currency, it may be a fifth, or a tenth, or a twentieth. A large command of resources in the form of currency renders their business easy and smooth, and puts them at an advantage in bargaining; but on the other hand it locks up in a barren form resources that might yield an income of gratification if invested, say, in extra furniture; or a money income, if invested in extra machinery or cattle. A man fixes the appropriate fraction after balancing one against another the advantages of a further ready command and the disadvantages of putting more of his resources into a form in which they yield him no direct income or other benefit."³

8. Hutchison, *op. cit.*, p. 73.

9. *Principles*, Book 4, chap. 7. G. J. Stigler, *Production and Distribution Theories*, pp. 102-3.

1. *Principles*, p. 430.

2. H. J. Davenport, *The Economics of Alfred Marshall* (Cornell University Press, 1935), p. 457.

3. J. M. Keynes: *A Treatise on Money*, I, 230.

The supply is determined, in the pure system, by the general equilibrium relations of the whole system; or more concretely, by the trading relations of a country with others, by the technical and cost conditions of production of money, and so on; and, with paper money, by decision of the political authority. Then, given the two — M and k in the equation $M = ky^4$ — the general level of prices is determined. The third main “tool” derived from the stationary state was Say’s Law,⁵ perhaps not stated explicitly or rigidly, but nevertheless essential for long-run normal equilibrium in Marshall’s sense.

With this analytical equipment Marshall had to face the facts of unemployment, crisis, and disturbances in the real world. But though he repeatedly warns us in the *Principles* that these factors must be kept in mind,⁶ and at the very end speaks of the “complex actions and reactions of credit, of foreign trade, and of modern developments of combination and monopoly”⁷ which must be looked at before practical conclusions are reached; and though he contemplated a systematic book on the subject, which was never written⁸ (*Money, Credit and Commerce* being for the most part a collection of early writing and evidence); despite all this his own views are not those of a man who is a disinterested tool user, but rather they are the reflection of a definite “vision” of reality; they are derived from a conceptual system which is supposed to represent fairly closely essential and universal forces acting in the world.

The kind of approach of a person holding Say’s Law is familiar. Thus for Marshall unemployment was a phenomenon of credit expansion and contraction. Some accidental factor — a good harvest, the discovery of new colonial territory — acts on people’s expectations; prices, wages, and profits rise as credit expands; eventually “distrust takes the place of confidence, failure and panic breed panic and failure. The commercial storm leaves its path strewn with ruin. When it is over there is a calm, but a dull heavy calm. Those who have saved themselves are in no mood to venture again.”⁹ Although “having the power to purchase,” men have not chosen to use it.¹ But the next day dawns fine; normal times are back again. In the same way, surveying the economic state of the nation for the benefit of

4. Marshall did not formulate the theory in this form.

5. *Principles*, p. 591.

6. *Ibid.*, e.g., at beginning of chaps. 1 and 8 in Book V.

7. *Ibid.*, p. 601.

8. Keynes, *Essays in Biography*, p. 161.

9. *Elements of the Economics of Industry* (London: Macmillan, 1892), pp.

152–53.

1. *Principles*, p. 591.

the Royal Commission on the Depression of Trade and Industry (1886),² he remarks that the evil of the era of low prices since 1873 had been exaggerated. Indeed, on the whole, low prices had "conducted more to solid progress and true happiness than the alternations of feverish activity and painful retrogression which have characterised every preceding decade of this century."³ The only theoretical reason for unemployment (apart from the normal switching of the labor force from old into new industries) lay in the lag between falling prices and falling costs. Some costs were fixed in money terms and so increased in real terms. There could conceivably be some "suspension of industry" for this reason; but probably not a great deal in this case, because prices were falling so slowly. But whatever the theoretical possibilities, unemployment was *not* increasing, simply as a matter of fact. "I believe that the statistical evidence brought forward to prove it is invalid."⁴ Frictional employment of the kind theoretically possible could be largely avoided, in any case, by the general adoption of the device of adjusting all contracts longer than a certain period, by an index of general prices.⁵

Just as Marshall applied this part of his central analysis fairly unreservedly to real problems, so with his theory of interest. The Gold and Silver Commissioners asked him what he thought would be the effect of an influx of gold into Lombard Street.

He answered, very much as Ricardo or even Smith would have done "... the supply of gold exercises no permanent influence over the rate of discount. The average rate of discount is permanently determined by the profitableness of business. All that the influx of gold does is to make a sort of ripple on the surface of the water."⁶ His reasoning seems to be: an inflow of gold to the money market lowers the interest rate temporarily; this encourages "speculative ventures"; credit is extended generally; but this raises prices. With higher prices, assuming no change in habits, people wish to hold more cash; this tightens the money market again, and the old rate of interest is restored.⁷

Marshall's concern for explanations of real phenomena ultimately unified by being related to the forces of long-run supply and demand, also led him to pass lightly over many relationships which he perceived clearly enough, but did not develop.

1. He noticed that "the demand for Fixed capital is liable to

2. *Official Papers by Alfred Marshall* (London, 1926).

3. *Ibid.*, p. 9.

4. *Official Papers: Evidence before Gold and Silver Commission*, p. 92.

5. *Ibid.*, p. 9.

6. *Ibid.*, p. 41.

7. *Ibid.*, pp. 38-44.

more extreme fluctuations than the demand for commodities that are wanted for immediate consumption, and the trades which make Fixed capital are more affected than any others by alterations of commercial prosperity and adversity." He gave the examples of pig iron and the building trade.⁸

2. He recognized some stability in consumption purchases "when the wave of prosperity is descending"; for "the statistics of total consumption of such commodities as tea, sugar, butter, wool, etc. prove that the total purchasing power of the people does not meanwhile fall very fast."⁹

3. As had many writers before him,¹ he saw how a cumulative process could be started by a divergence between the "nominal" and the "real" rate of interest, due to a changing level of prices. This, however, was "most conspicuous in the market for short loans — a market which differs in many of its incidents from any other, and a full discussion of their influences must be deferred."² As with so many others, the promised "full discussion" never eventuated.

4. He clearly distinguished the motives of individuals which lay behind k in the quantity equation, from the facts of business organization affecting velocity of circulation in a more mechanical way.³ Normally, there is apparently no hoarding for reasons of speculation and security; but there *may be*, if a change in the value of the currency is expected, e.g., appreciation against silver if a bimetallic system cannot be held.⁴ Later, in *Money Credit and Commerce*, as Hansen points out,⁵ he supposes that assets, and not only incomes, are relevant. Thus he says: "Let us suppose that the inhabitants of a country, taken one with another . . . find it just worth their while to keep by them on the average ready purchasing power to the extent of a tenth part of their annual income, together with a fiftieth part of their property . . ."⁶

Thus we have $M = kY + kA$, where A stands for total assets. This is not far from Keynes's "liquidity preference" motive, since security prices, i.e., the rate of interest, obviously affect the total value of assets owned.⁷ However, it is significant that Marshall made no use of this idea. If he had thought it through he must inevitably

8. A. H. Hansen: *Business Cycles and National Income*, pp. 273-74.

9. *Principles*, p. 92.

1. E.g., Ricardo and Thornton.

2. *Principles*, p. 493.

3. *Official Papers*, p. 35.

4. *Ibid.*, p. 59.

5. *Money Credit & Commerce* (London, 1923).

6. *Ibid.*, p. 44.

7. This point is made by A. H. Hansen, *Monetary Theory and Fiscal Policy*,

have come to see the connection between the rate of interest and the quantity of money. Indeed, his reasoning was not satisfactory on this very matter, when rather hard pressed at one stage by the Gold and Silver Commissioners. Why, they asked him, should an increase of gold in Lombard Street cause a "flutter" upwards in security prices, speculation, and increased general prices, when for the past two years, there had been a great "plethora" of capital seeking discount on the market? Why, it was implied, would the extra money not simply remain on the market, uninvested?⁸ Again, Marshall's preconceptions were against such a possibility. If he had admitted the possibility that money may permanently affect the rate of interest, he would have been committed to overhaul his whole system of long-run equilibrium.

III. ASPECTS OF ECONOMIC DEVELOPMENT

As in his theory of money, in his theory of development Marshall saw many significant relationships, but because of his concern for unity, failed to develop them clearly.

A. Increasing Returns

Possibly the most obvious and best known example is his discussion of increasing returns. The economies of large-scale production are emphasized strongly. A large firm has advantages in the use of material,⁹ in the use of specialized machinery,¹ in the invention of new machinery,² in buying and selling on a large scale,³ in the use of (especially managerial) specialized skills,⁴ and in finances.⁵ If so, how does competition exist at all?

Marshall gave several reasons. First, there was the doctrine of "external economies." There were two main categories of these — those internal to the industry — "Economies in the use of specialised skill and machinery . . . which depend on the aggregate volume of production in the neighbourhood" — and others external to it — "those connected with the growth of knowledge and the progress of the arts," depending chiefly "on the aggregate volume of production in the whole civilised world."⁶ But, as Sraffa⁷ pointed out, the latter

8. *Official Papers*, pp. 49–51.

9. *Principles*, p. 232.

1. *Ibid.*, p. 3.

2. *Ibid.*, p. 4.

3. *Ibid.*, p. 5.

4. *Ibid.*, p. 6.

5. *Ibid.*, p. 7.

6. *Ibid.*, p. 220.

7. Sraffa, "The Laws of Returns under Competitive Conditions," *Economic Journal*, XXXVI (Dec. 1926).

were not really applicable to Marshall's "particular equilibrium" approach to industrial analysis, and the former are of little importance as long as they are supposed strictly external to particular firms. The whole construction was a clumsy one, in so far as it was meant to fit into the theory of normal long-run equilibrium. It was a fruitful idea⁸ forced into a system where it did not really belong.⁹

The second reason Marshall gave to explain the absence of widespread monopoly was the existence of special markets. "But many commodities with regard to which the tendency to increasing returns acts strongly are, more or less, specialities; some of them aim at creating a new want, or meeting an old want in a new way. Some of them are adapted to special tastes, and can never have a very large market; and some have merits that are not easily tested and must win their way into general favour slowly. In such cases the sales of each business are limited, more or less according to the circumstances, to the particular market which it has slowly and expensively acquired; and though the production itself might be economically increased very fast, the sale could not."¹ This, of course, is the solution of Sraffa, Joan Robinson, Chamberlin and the others. If Marshall's writing is perused carefully enough, almost every conclusion of interest that these and other writers came to, can be found more or less clearly stated. A notable example is the unpublished paper on retail prices,² where there is a clear description of what Chamberlin later called "product differentiation," among other things.

Yet Marshall did not develop a specific theory along these lines. He left the matter curiously incomplete. For in trying to find how increasing returns could be made compatible with normal competitive equilibrium, he had arrived at a reason which was itself incompatible. So, while keeping "special markets," he minimized their importance. Monopolistic prices were like "snow flakes which rise as they fly past a house in a strong wind, not because gravitation is in abeyance, but because it is overborn by the force of wind eddies."³

The third and best known of Marshall's reasons for the existence of competition lay in "the balancing of the force of life and decay"

8. E.g., M. Fleming, "External Economies and the Doctrine of Balanced Growth," *Economic Journal*, LXV (June 1955), and the literature referred to there.

9. Thus Stigler says (*op. cit.*, p. 76) "As a tool in the explanation of economic history, the doctrine of external economies (but in a different form than proposed) seems to have considerable serviceability. As a device for the elucidation of relative prices, it seems to have a very restricted scope."

1. *Principles*, p. 239.

2. *Memorials*, p. 354.

3. *Ibid.*, p. 356.

in the firm; in the mortality of the businessman and the inferiority of his descendants.⁴ Thus at any time the industry consists of many firms, some rising, some falling. Taking a broad view "we need not trouble ourselves with these eddies on the surface of the great tide";⁵ and if we wish to look closer, it is enough to consider a "representative" firm. But the analogy, while useful, was of doubtful validity, if carried too far. Marshall himself came to recognize this: "every recent decade has contained some episodes which suggest that it [i.e., the idea of decay in the management of the firm] may probably be greatly changed, either in substance, or in the methods by which new life is brought into old bodies."⁶

Speaking generally of Marshall's treatment of increasing returns, we can say then that the idea of itself was of considerable importance. But instead of being followed along its own peculiar paths, it was limited and fitted into the general system of stationary equilibrium. In the course of limiting it, new discoveries were made — the concept of "external" and "internal" economies, the ideas of the "particular market" and of organic growth and decay. Each of these deserved study in its own right; instead (uncomfortable) niches were also found for them in the general scheme.

B. "Organization"

"The central unity of action between the laws of nature in the physical and the moral world," Marshall wrote, "is that the development of the organism, whether social or physical, involves an increasing subdivision of functions between its separate parts on the one hand, and on the other a more intimate connection between them."⁷ "Organization" is the structure of the organism thus adapted to its environment; it is in the main spontaneous and natural, but not incapable of modification by the purposive action of man.⁸ In the more special sense in which Marshall uses it "it has many forms, e.g., that of a single business, that of various businesses in the same trade, that of various trades relatively to one another, and that of the State providing security for all and help for many."⁹

Used in this way, the concept of "organization" is obviously a very broad one, and mainly useful for the elucidation of historical development. Yet Marshall puts it on a par with land, labor, and

4. *Ibid.*, p. 318.

5. *Principles*, p. 314.

6. *Industry and Trade* (2d ed.; London, 1919), p. 316.

7. *Principles*, p. 201.

8. *Ibid.*, p. 206.

9. *Ibid.*, p. 115.

capital as a "factor of production." How can it possibly be integrated into the general theory of Book V, if it is left so vague? This is all the more remarkable, as Marshall is so aware of the roughness and approximate nature of the other concepts. But he explains that, while "the abstract and mathematical point of view," which would assimilate all factors to a common denominator (so that, for example, the return to labor is rent of labor service plus interest on capital), is, on its own ground "incontestable"; yet for ordinary analysis we do away with "the incessant enumeration of details of secondary importance" by using broader categories which are roughly correct¹ (e.g., labor is supposed to consist more in pure "labor service" than the service of capital invested). Thus he is clear about the connection in these cases, and even takes some trouble to explain it; yet "organization" is left hanging in mid-air. Possibly there is a connection in that the entrepreneur "organizes" his business; and this activity may in some sense be considered a separate factor of production; but the resemblance between this and the wide ranging discussion of Book IV is not close, to say the least. The conclusion must be that this is another example of Marshall's trying to force a separate consideration into a mold not its own.

C. "Substitution"

In the static "normal" theory the principle of substitution requires the assumption of a given "technological horizon," as Schumpeter would say. Substitution must be between existing known factors of production, on one side; and existing, known items of consumption, on the other. Inventions leading to new factors, and new products, are incompatible with stationary equilibrium in the marginalist sense.

Despite this, substitution in Marshall is conceived as a dynamic process involving new techniques and new products. It is true that in defining the term in Book V he adds the essential qualification "as far as the knowledge and the business enterprise of the producers' reach,"² but this is not the impression we get from the rest of the book. Thus he speaks of the function "which Roscher calls a characteristic task of the modern manufacturer, that of creating new wants by showing people something they never thought of having before";³ and of the evolution of new methods and the use of new machinery. Indeed, his substitution principle is surely closer to the

1. *Ibid.*, p. 649.

2. *Ibid.*, pp. 234, 284.

3. *Ibid.*, pp. 234, 284.

Darwinian struggle for survival and linear evolution, than to static mechanics. Yet in the presentation in the *Principles* the former kind is, with superficial ease, included in the latter.

D. "Character" and "Activities"

Schumpeter said once, in reference to Marshall, "I confess that few things are so irritating to me as the preaching of mid-Victorian morality, seasoned by Benthamism — the preaching from a schema of middle-class values that knows no glamour and no passion."⁴ And, of course, many others today are irritated (as in Marshall's own time most people were pleased) by the constant references to the "average level of human nature"; to the virtues of steadfastness, enterprise, nobility of purpose, and so on. Yet, as Talcott Parsons has shown, this element in Marshall's thought is a systematic one, and of great importance.⁵ The quality of character which Marshall admires so strongly is the essential force which raises man over the animals, and civilized man over the "more ignorant and phlegmatic of races."⁶ Those whose "mental horizon is wider, and who have more firmness and elasticity of character,"⁷ prize work for its own sake. Indeed, "Work, in its best sense, the healthy energetic exercise of faculties, is the aim of life, is life itself."⁸ Social evolution seems to occur something like this: higher income stimulates new and "higher" wants; this in turn improves character; the higher average level of character brings forth labor more efficient and of higher quality; moreover the quality of more and more vividly "realizing the future" — projecting more wants into the future — increases. This means a greater supply of saving, increased division of labor and use of machinery; which in turn leads to a higher level of income and activities; character further improved, and so on.

This emphasis on work, or more generally, "activities," is, as Parson points out, scarcely compatible with the utilitarian presuppositions of neoclassical value theory. "What is the aim of life, what is life itself, cannot well be interpreted as a cost which must be incurred in the attainment of ends outside life itself."⁹

Here then, we have another element in Marshall's thought which does not fit easily into his central thesis. Indeed, it not only does

4. Schumpeter: *Ten Great Economists*, p. 104.

5. Talcott Parsons: *The Structure of Social Action*, chap. 4.

6. *Principles*, p. 439.

7. *Memorials*, p. 115.

8. Parsons, *op. cit.*, p. 148.

9. *Ibid.*, p. 148.

not fit easily, but in some ways appears to be a more general principle than the one which is supposed to embrace it. For example, Marshall is careful to point out that his discussion of consumption is "an elementary analysis of an almost purely formal kind";¹ and it is evident that anything beyond this is based directly on the relation of wants to activities. Moreover, in Marshall, this position is more than a simple case of "philosophical" preconception; it is a generalization which could be tested and, as Parsons has shown, enters even into his detailed explanation of the responsiveness of the supplies of labor and savings to changes in the reward for them.²

IV. CONCLUSION

Keynes accused the "Classical Economists" of making a false division between the theory of value and distribution on the one hand, and the theory of money on the other. "We have all of us become used to finding ourselves sometimes on the one side of the moon and sometimes on the other, without knowing what route or journey connected them, related, apparently, after the fashion of our waking and dreaming lives."³

It is obvious that the very opposite is true of one classical economist at least, Alfred Marshall. We have seen that he approached monetary problems with three tools which were all closely integrated with each other, and with his central concept of "normal equilibrium." These were the concepts of the demand and supply of loanable funds, the concept of the demand and supply of money, and Say's Law of markets. It is apparent from the discussion that these tended to be used not simply as tools, but as though they were a fairly realistic model of reality. Moreover, their central position in Marshall's thought led to a number of promising approaches not being followed up.

Secondly, we have seen how a number of other concepts — increasing returns, the ideas of "substitution" and "organization," and the theory of activities — which really needed separate treatment, were instead rather out-of-place portions of the structure which supported Book V of the *Principles*.

It is remarkable that many of the least comfortable of these ideas, given free play, would tend to weaken the structure as a whole. Yet we cannot accuse Marshall of making Mill's mistake, and supposing that his words were the last to be said on the subject; he spoke

1. *Principles*, p. 72.

2. Parsons, *op. cit.*, pp. 141-50.

3. Keynes, *General Theory*, p. 292.

of "the far greater work" for which his own generation had only prepared the way.⁴ Nor can we accuse him, least of all, of intellectual dishonesty. What, then, is the explanation? The answer is, I think, that Marshall was one of those great economists with a *vision* of the social process, which is in the end incapable of proof, as it is of disproof. Marshall's vision was of "the deep silent strong stream of normal distribution and exchange";⁵ and this was one of those ideas, which, in his own words, "can never die but are an existing yeast ceaselessly working in the Cosmos."⁶ Marshall thought it important that other ideas should be related to this central vision; but none of them could, almost by definition, displace it. They had to be subservient to it. The force of gravity would prevail in the end. The question then arises: how many exceptions can be made? How far must the fundamental force work below the surface of things before we can ignore it? How we answer this question largely depends on the particular circumstances of the time, or the recent past.⁷

THE AUSTRALIAN NATIONAL UNIVERSITY

4. *Memorials*, "The Old Generation of Economists and the New," p. 310.

5. *Principles*, p. 522.

6. *Memorials*, p. 374.

7. E.g., see N. Kaldor, "A Model of Economic Growth," *Economic Journal*, LXVII (Dec. 1957), 595-96.

CULBERTSON ON INTEREST STRUCTURE: COMMENT

By L. S. WEHRLE

Although current theory of the maturity structure of interest rates is in no sense complete or integrated, it is a far more acceptable body of doctrine than is indicated in John Culbertson's recent article.¹ It is true that since the theory exists only among the pages of scattered journal articles it is extremely difficult for one to form a correct perspective of its current status. Culbertson's article, while informative and interesting in its description of institutional demand for securities, does little to lessen this difficulty since it presents a description of the current theory that is both incomplete and, in places, incorrect.

This note will set forth an interpretation of existing theory and will discuss Culbertson's theoretical treatment of expectations.² Both because of its development and for expositional convenience, the theory will be discussed under two headings: expectations and uncertainty. The discussion of expectations refers to the influence on present interest rates of expected future rates while the discussion of uncertainty refers to the influence on present rates of uncertainty of future rates. The term expectations refers to a measure of central tendency and the terms uncertainty and risk both refer to measures of dispersion of an investor's subjective probability distribution of anticipated outcomes.

Irving Fisher's statement concerning expectations³ — the present long rate is an average of expected short rates discounted over the life of the longest security — assumes unanimous and "certain" expectations of future interest rates. The current version relating to expectations, developed by John Hicks,⁴ Nicholas Kaldor,⁵ Michal

1. "The Term Structure of Interest Rates," this *Journal*, LXXI (Nov. 1957).

2. This note is necessarily a very brief sketch of current theory. No attempt is made to document the interpretation given. The author is indebted to Professor James Tobin for discussion of this note as well as for extensive guidance in the study of this subject. Thanks are also due to Richard Porter for helpful discussion.

3. "Appreciation and Interest," *Publications of the American Economic Association*, Vol. 11 (August, 1896); especially pp. 23-29 and 88-92.

4. *Value and Capital*, especially pp. 163-66.

5. "Speculation and Economic Stability," *Review of Economic Studies*, VII (Oct. 1939).

Kalecki,⁶ and Friedrich Lutz,⁷ differs from the Fisher version primarily in its introduction of differences among investors concerning: (1) expectation of future interest rates, (2) holding or planning periods used, and (3) the amount of investable funds held. These differences are illustrated in Lutz's description of the investment decision.

He will discount the price at which he expects to sell the bond at the date when he wants to disinvest (this price is dependent on what he anticipates the long rate will be at that date) and all the interest payments up to that time; using as the discount factor for each year the short rate which he expects to prevail in that year.⁸

Because of these differences among investors, there is no simple sense in which it can be said that the rate structure indicates the underlying investor expectations.

A further characteristic of investors, the degree to which they are uncertain about future rates constitutes the basis of the theory of a number of writers. Although these writers all view the investor as uncertain about future interest rates, they differ in their view of the investor's degree of certainty regarding his future liquidity needs.

One group, including R. F. Kahn,⁹ Mrs. Joan Robinson,¹ and Friedrich Lutz,² assumes the investor to be certain of the time path of his future liquidity needs and to minimize the risk of capital loss. The possibility of future changes in interest rates gives rise to risks of loss of capital value to any investor who must sell a security before it matures. Among securities of different maturities, each will have a different future date when capital-value risk is reduced to zero — its date of redemption. Thus a person who knows his future liquidity needs can minimize — indeed, in principle, eliminate — capital-value risk by purchasing maturities that match the time path of his future liquidity needs. The essence of the theory is that investors desire to minimize risk by purchasing appropriate maturities, even at the sacrifice of possible expected return. The resulting rate structure is determined by the relative supplies and demands for maturities and may be of any configuration.

A corollary of this view is that the capital-value risk of a security exists only relative to the particular investor considering the security. The risk is not embodied in the security itself.

6. *Theory of Economic Dynamics* (London, 1954).

7. "The Structure of Interest Rates," this *Journal*, LV (Nov. 1940).

8. *Ibid.*, p. 49.

9. "Liquidity Preference," *Manchester School*; April 1954.

1. *The Rate of Interest* (London, 1951), first essay.

2. *Op. cit.*

In fact, the investor is not so limited in his choice of maturities as the above view implies. First, the investor may still maintain a risk free portfolio by purchasing maturities less than as well as equal to the time-availability of his funds, since securities that mature before the funds are needed may be reinvested or held as cash without risk of capital loss until needed. This broadens the range of maturities available to this investor though it introduces uncertainty concerning the interest return that will be realized over the period. Second, the investor may choose not to minimize capital-value risk if expected return on alternative maturities is sufficiently attractive.

It is possible in the case just discussed for investors to choose portfolios free of capital-value risk; this is not possible if the future liquidity needs of the investor are assumed uncertain. The implications of this assumption, developed by John Hicks,³ Harry Markowitz,⁴ and James Tobin,⁵ are that the investor is not interested in purchasing a particular pattern of maturities since he is unable to predict his future liquidity needs. Although he is unconcerned with maturity *in this sense*, he is not necessarily indifferent to the greater capital-value risk associated with long maturities. If the investor is assumed to "dislike" risk — to have a declining marginal utility of wealth function — he will hold long securities only if the interest differential is sufficient to compensate for the additional risk incurred. This will cause the rate structure to be an increasing function of maturity if expectations are not considered.

The theory can be summarized by noting two polar types of investors. First, there are those investors, class (1), who know their future liquidity needs and minimize the associated capital-value risk by purchasing appropriate maturities. As previously stated, appropriate refers to maturities less than as well as equal to the time-distance of future liquidity needs. Second, there are investors, class (2), who are uncertain of their future liquidity needs and thus choose their portfolio by consideration of risk and return. Their sensitivity to future rate expectations depends on the degree to which they are willing to incur risk in order to increase expected return. Indeed, both *may* choose their portfolios on the basis of risk and return; the difference between them is that one *can* choose a riskless portfolio while the other *cannot*. For emphasizing the importance of future liquidity needs in relation to interest structure we are indebted to

3. *Op. cit.*

4. "Portfolio Selection," *The Journal of Finance*, Vol. 7 (Mar. 1952).

5. "Liquidity Preference as Behavior toward Risk," *Review of Economic Studies*, XXV (Feb. 1958).

Culbertson (pp. 492-93). Previously this distinction has not been carefully recognized in much of the literature.

The following section will be concerned with two aspects of Culbertson's analysis: his critique of what he terms the expectational theory and the theoretical basis for his "nonspeculative" method of portfolio selection.

First, his presentation of the expectational theory is unnecessarily narrow. It is not true that this theory is "based upon the theoretical consideration of the implications of confidently held expectations" (p. 487), that it depends upon a "planning period [equal] . . . to the longest investment asset being considered" (p. 497), or that it assumes "the expectations of individual speculators are . . . ordinarily unanimous or self-consistent" (p. 496). Furthermore, his conclusions that "'market expectations' can [not] logically be inferred from the actual structure of interest rates" (p. 490) and "that speculation in debt markets will be primarily based upon short planning periods" (p. 497) have been clearly stated by Kahn, Kaldor, and Lutz.

Second, in relation to the relevance of expectations, Culbertson suggests that the majority of investors follow a "nonspeculative" method of portfolio choice which is defined as choice not based upon expectations of future rates. But to the contrary, expectations are not so easily banished from an investor's decision process, at least not if the term expectations is correctly interpreted.

Before proceeding to show why this is true, it is important to understand what is meant by the term liquidity. Culbertson defines it as the "ability to be turned into cash on short notice on definite and favorable terms" (p. 491). The first element refers to marketability and the second to capital-value risk. Primarily, he refers to capital-value risk; "debt of shorter maturity is the more liquid" (p. 491). Thus his use of liquidity is equivalent to capital-certainty which, in turn, is the obverse of capital-value risk. It seems preferable to use the distinct terms.

There are two methods of obtaining capital-certainty according to Culbertson: the class (1) investor, as described in the preceding discussion of uncertainty, can hold specific maturities while the class (2) investor can hold short maturities (pp. 492-93). Each of these suggestions will be considered.

Concerning the class (1) investor three points should be noted about his "nonspeculative" portfolio — "select a portfolio maturity structure suited to the liquidity needs of the investor . . . [by] making choices on some basis that is independent of any particular expectations regarding the future course of . . . interest rates" (p. 499).

First, there is no particular reason to assume this investor to desire perfect capital-certainty if he could obtain a higher return by taking some risk; the degree to which he chooses capital-certainty depends both on his preferences for risk and return and on the risk and return differentials available. Second, even if the investor desires complete capital-certainty, he will still be free to choose among a range of maturities which are equal to or less than the time-availability of his funds. Within this range, expectations of rate movements could govern asset choice in the manner described by the expectational theory. Third, the assertion that capital-certainty demand results in a positive correlation between yield and maturity ("most investors must be sensitive to liquidity considerations," p. 492) gives insufficient recognition to an opposite incentive that causes an aversion to short-term assets. This is the desire for obtaining the certainty of a known interest return over a long period of time which Mrs. Robinson calls "income-certainty." An investor (e.g., a pension fund or life insurance company) who contracts today to make specific future deliveries of dollars and hence is interested in knowing today what interest income will be earned in the interim, would be primarily interested in the greater income-certainty provided by long-term securities. In relation to this class of investor, the conclusion reached is that Culbertson's "nonspeculative" method of portfolio selection will apply only to investors who desire capital-certainty and income-certainty concurrently. If only capital-certainty is desired, then, within a range of maturities *less than* or equal to the time-availability of the funds, assets may be chosen to maximize the expected return of the portfolio. Likewise, if only income-certainty is desired, then, within a range of maturities *greater than* or equal to the time-availability of the funds, assets may be chosen to maximize the expected return of the portfolio. Finally, if maximization of expected return is the sole criterion, no constraints are placed on the choice of maturities.

While the "nonspeculative" method of asset choice might describe, at least in part, the behavior of the investor with known liquidity needs, for the class (2) investor expectations of future rates must be involved in the decision process. Unless the investor chooses only capital-certain assets such as cash and Series E Bonds, he must choose among assets of differing degrees of risk. This can be seen by an examination of four elements of an investment decision.

(a) Estimates of return — the expectations of the investor regarding future interest rates on different securities.

(b) Estimates of risk — an estimate by the investor of capital-value risk inherent in the different market assets; i.e., the anticipated degree of price variability of each asset.

(c) Preferences — assume the investor to choose assets on the basis of their capital-value risk (capital-certainty) and expected return; thus there is a risk/return preference.

(d) Market prices — the investor is presented with a constellation of prices and yields in the market.

The investor is assumed to choose that combination of assets which, given the objective market prices (d) and his subjective risk and return estimates (a and b), will optimize his position in relation to his preferences (c).⁶

In relation to the class (2) investor, it is stated that most investors choose capital-certain assets whose selection is "not . . . based upon any specific expectations regarding the future debt prices or market conditions, but rather on the general characteristics [read short maturity] of the asset" (p. 491). However if the view presented above is accepted, the choice of short-term assets is a result of expectations as well as a means of avoiding them. For if the investor chooses between return and capital-certainty, the choice necessarily involves an estimate of the return and capital-value risk on available assets; this, in turn, necessarily involves an estimate of future interest rate movements. It is important to distinguish between an investor's estimates of risk and his preferences for risk.

The implications of this use of expectations may be seen more clearly by examining Culbertson's hypothesis that "the wide differential between short-term and long-term rates [in the thirties] can be interpreted as consisting mainly of an unusual liquidity premium associated with the . . . [increased] demands for liquidity in investment assets" (p. 513) and the . . . "drop in the supply of short-term debt" (p. 510). This is a possible, but not a necessary explanation. Alternatively, investors during this period, due to the state of flux and uncertainty, may have increased their estimates of the "risk content" of all maturities. The possibility of large fluctuations of interest rates and consequently large fluctuations in the price of long-term debt was regarded as more likely than it had been previously. Even with unchanged preferences an investor would retreat from long maturities. Thus the price of capital-certainty (the difference between the short and long rates) is increased until the risk/return relation is again reduced to desired levels. According to this hypothesis, changes in rate differentials may be explained by changes in future outlook without postulating shifts in the investor's preferences.

YALE UNIVERSITY

6. Cf. James Tobin, *op. cit.*, for a formal discussion of this method of viewing the investment decision.

REPLY

By J. M. CULBERTSON

My article did not give a general description or critical review of the recent literature on the interest rate structure. Its objective was to pull together a realistic theory and to apply it to explanation of actual interest rate behavior. The literature dealing with the theory of the rate structure, and with points that are involved in the theory of the rate structure, is large, capable, inconsistent, and in some cases self-contradictory. To review it carefully and attempt to give an integrated interpretation of the terminological, methodological, and behavioral questions involved would require considerable space, and would presuppose some understood framework upon which the criticism could be based. A cursory treatment such as Wehrle's that does not disclose the differences in the literature and the issues seems to me misleading rather than helpful.

The adequacy of the existing literature is necessarily a matter of interpretation and point of view. It seems to me that the continued prevalence of the objectionable expectational theory¹ (side by side in some cases with other doctrines not recognized as inconsistent with it) and of interpretations of past economic events and of policy conclusions based upon it, as well as the internal problems in the existing literature, argue that there is a need for new work.²

Wehrle's charge that my "description of the current theory . . .

1. I defined "the expectational theory" specifically as "the theory [that] argues that the interest rate on a long-term debt tends to equal the average of short-term rates expected over the duration of the long-term debt" (p. 487). I indicated that this is not the only interpretation of the influence of expectations on the rate structure and pointed out the most important alternative interpretation (p. 487, footnote 5). I took exception to the expectational theory and gave an analysis of the terms of reference of debt speculation that is inconsistent with it, but I did not, as Wehrle implies, offer any derogatory summary of all recent theoretical writings in this area.

2. It is perhaps significant that Wehrle's necessarily brief "interpretation of existing theory" says nothing about how the interest rate structure might be expected to behave, consisting entirely of description of a way in which the investment decisions of certain hypothetical investors might be described.

More broadly, some of the literature on the rate structure that is unobjectionable on theoretical grounds is correspondingly unhelpful with real problems in that it is devoid of explanatory content, consisting essentially of deductive exercises as to what would happen if investors behaved in certain often improbable ways. For theories emphasizing expectations, the process of giving real explanatory content to a theory is that of deciding the nature of the necessarily incomplete and imperfect expectations actually used in decision-making and the manner in which they are used, of asserting the predominance of certain patterns of expectationally governed behavior from among the innumerable conceivable ones.

is . . . in places, incorrect" (p. 601) is anomalous in that most of the "description" to which he apparently has reference is not supplied by me but represents his misuse of quotations. To compound the confusion, the "description" that he condemns is not, in fact, without its applicability to current doctrines. What seems to be in question is the interpretation of the expectational theory. Let me make several points regarding it:

(1) While I said that the expectational theory (as I defined it) is "the doctrine on the term structure of rates most influential recently among English and American theorists" (p. 487), I surely did not imply that it is universally held or that no one has objected to it.

(2) Of the three quotations that Wehrle offers to support his charge that my "presentation of the expectational theory is unnecessarily narrow" (p. 604), the first is slightly inaccurate³ and the other two are passages lifted from a general discussion of the influence of speculative behavior on the term structure. It is misleading to characterize the latter two as a "presentation of the expectational theory."

(3) My discussion of the role of speculation does, I believe, provide tools for critical consideration of alternative viewpoints and applies to the expectational theory. Its applicability in this connection does not seem to be understood by Wehrle. For example, when the distinction between maturity yields and "holding-period yields"

3. I said that the expectational theory "was based upon the theoretical consideration of the implications of confidently held expectations . . ." (p. 487). Wehrle supplies an "is" in place of the initial "was," which pushes the meaning in the direction of a characterization of the justification rather than the derivation of the doctrine.

That the theory derives from consideration of the implications of abstractly perfect expectations is a simple matter of record. It surely was not developed from any empirical study of the performance of interest rates or the characteristic behavior of investors. Irving Fisher's statement of his conception of the long-term rate as an average of the hypothetical short-term rates applicable to the various segments of the longer period in *The Theory of Interest* (pp. 313-14) was within a framework of assumed perfect foresight. He also said: "It is impossible to give a concrete example of a rate of interest for a long-term loan as an average of the year to year rates, because, as already noted, the year to year rates have only a hypothetical existence" (*ibid.*). In the portion of the same work dealing with the actual behavior of the rate structure, the author made no use of this relationship (chap. IX). Hicks's characterization in *Value and Capital* of the long-term rate as a series of "forward short rates" (pp. 144-47) belongs to a section based upon very restrictive assumptions (see pp. 124-27) and the author claims little for their realism (see pp. 126, 140). Lutz works out his theory on the assumption that "everybody concerned knows what future short-term rates will be . . ." before going on to apply it to analysis of actual rates (this *Journal*, LV (Nov. 1940), 36-37). What is distinctive about the Lutz article is that it does go on and applies to explanation of the actual rate structure what had been treated as an interesting hypothetical relationship among interest rates.

is clearly drawn, it becomes evident that the expectational theory does necessarily imply a planning period equal to the term to maturity of the longest debt being considered (although this is not consistently recognized by Lutz). If the comparison of prospective debt returns includes the yield to maturity on a long-term debt (in comparison with an average of short-term rates), then the period involved must be the period to maturity on the long-term debt, because the long-term "yield" measures the rate of return over only this time span. The rate of return over any shorter period would be some "holding-period yield" dependent upon the price of the long-term debt at the end of the period.⁴ If the planning period were a short one, this holding-period yield on the long-term debt would bear no necessary relationship to the initial maturity yield of the long-term debt but would depend mainly on the change during the period in bond prices and long-term interest rates (see my charts, p. 506).

Therefore the passage from Lutz quoted by Wehrle on page 602 and his statement that Lutz, among others, "clearly stated" that speculation in debt markets will be primarily based upon short planning periods (p. 604) are both inconsistent with the expectational theory.⁵ These observations thus cannot be used to support the adequacy of current doctrines including the expectational theory.⁶

Wehrle's major objection to the substance of my article has to do

4. It seems quite clear in the Fisher-Hicks conception of the long-term rate as an average of present and "forward" short-term rates that it is the duration of the long-term debt that determines the time span of the comparison, and this is recognized by Lutz in the passage quoted on p. 602. The necessity of using what I called "holding-period yields" when the planning period is shorter than the duration of the long-term debt is indicated by Lutz in the passage quoted by Wehrle in another connection (p. 602), but the idea is not incorporated into the substance of his theory.

5. The former in that it implies the necessity for a concept of holding-period yields while no such concept is used in the expectational theory. The relevance of the latter, of course, is that if the planning period is equal to the term to maturity of the long-term debt it is not short.

6. When Wehrle points out (p. 604) that the idea that market expectations cannot logically be inferred from the actual structure of rates is not a new one, he is, of course, right. But if he is suggesting, as might be inferred, that this idea is generally accepted or that it is a part of the expectational theory (as might be deduced from the fact that he associates it with Lutz), then he is wrong. Lutz says: "Although the fact that different people in the market hold different opinions about the course of the rates that may be expected to prevail in the future means that there is no precise sense in which we can call the long rate an average of the expected future short rates, it remains nevertheless true that the long rate (or bond yield) is, in the complicated way described in Section IV, the outcome of the whole pattern of expectations of the members of the market as to the future short rates during the time the bond has to run" (pp. 55-56). In attempting to "verify" the theory, Lutz essentially deduces from the rate structure existing at selected points in time what the "expectations of the market" should have been

with my argument that debt investment can be in a meaningful sense nonspeculative. He asserts that for investors "... who are uncertain of their future liquidity needs" ... "expectations of future rates must be involved in the decision process."⁷ He believes that he establishes this by the use of a particular analysis of the investment decision. It seems to me that what he offers is only another way of chopping up an idealized investment decision, which can prove nothing as to the way in which actual investment decisions are made. However, his characterization of the investment decision actually seems to point to the need of a concept of a nonspeculative investor. As it stands, it involves an inconsistency.

The investment decision, according to Wehrle,⁸ can be broken into four elements: "a. Estimates of return — the expectations of the

on the basis of his theory, and argues that these deduced expectations are consistent with the surrounding circumstances (pp. 55–59). The deduced expectations are complete enough to include the approximate time pattern of expected future changes in short-term rates (p. 57).

For an illustration of a conflicting approach to deducing interest rate expectations from the rate structure, one based on assumed speculative operations depending upon near-term expectations regarding long-term rates, see Roland I. Robinson, *The Management of Bank Funds*, pp. 219–30.

7. Pp. 603, 605. By investors who "are uncertain of their future liquidity needs" I assume that he means those who are uncertain as to the time at which they will be forced to or desire to liquidate their investment, that is, as to their future cash needs. In the usual terminology, probabilities regarding future cash needs would determine a definite estimated liquidity need.

Wehrle is incorrect in attributing to me the view that the investor who does not know the timing of his cash needs will necessarily hold short-term securities (pp. 604, 606. That will depend upon the time pattern of probabilities of cash need. "For some ... the proportion of investment assets that needs to be in liquid form is very small" (p. 492).

Wehrle's discussion of nonspeculative behavior in the case of the investor with *definite future cash needs* (pp. 604–5) seems to misrepresent my position in two respects. I did not, as is implied, make desire for liquidity an absolute and overriding consideration in all cases, for example, I said, "... most investors must be sensitive to liquidity considerations when considering the disposition of a significant portion of their debt-invested funds" (p. 492). And I did not take the position that all investors prefer short-term debt, but made the marginal preference for short-term debt in our economy depend upon institutions and preferences (pp. 492–93). Wehrle also argues that an investor who knows when he will need his cash, even if he "desires complete capital-certainty ... will still be free to choose among a range of maturities which are equal to or less than the time-availability of his funds. Within this range, expectations of rate movements could govern asset choice in the manner described by the expectational theory" (p. 605). But this obviously would be speculative behavior in my meaning of the term, and would be inconsistent with "income certainty." Investors who insist on "capital certainty" but are indifferent to "income certainty" are probably rare.

8. Apparently this interpretation of the investment decision is based upon the cited article by James Tobin in the *Review of Economic Studies*. This issue of the *Review* has not yet made its way to Wisconsin and I have not seen this article.

investor regarding future interest rates on different securities. b. Estimates of risk — an estimate by the investor of capital-value risk inherent in different market assets; i.e., the anticipated degree of price variability of each asset. c. . . . risk/return preference. d. Market prices” . . . “the choice of short-term assets is a result of expectations as well as a means of avoiding them. For if the investor chooses between return and capital-certainty, the choice necessarily involves an estimate of the return and capital-value risk on available assets; this, in turn, necessarily involves an estimate of future interest rate movements” (pp. 605–6).

This makes it appear that an investor would say, for example: “I am buying long-term bonds because I am confident that the price immediately is going to rise, and they will have a high return, but I must recognize that I am subject to capital-value risk because prices of bonds are variable.” But this is absurd. This investor is *counting on* the bond price fluctuating, in a particular way. He will be disappointed if it does not. The *risk* of the investor who bases his purchase upon specific expectations obviously is the risk that the expectations will prove to be incorrect. His estimate of the liquidity of various investment assets would also be based upon his specific price expectations, although liquidity probably is of relatively little importance in the planning of active speculators. A “liquid” asset to the speculator would be one the price of which was expected surely to be continuously stable or rising during the planning period; if it were expected that the price of the asset might be depressed at some time during the planning period, it would be a less “liquid” asset. What Wehrle depicts is an investor who is half fish and half fowl. He is a speculator in thinking about returns, a nonspeculator in thinking about probable interim terms of liquidation. Such behavior is as improbable as it is irrational.

To make sense of the elements that Wehrle has mixed together, it is necessary to distinguish the speculative and the nonspeculative investor. The speculator has definite expectations and plans consistently on the basis of them as indicated above. The nonspeculator chooses not to base his operations upon specific interest rate expectations, and he consistently avoids doing so. He selects a portfolio of debt that on the basis of the characteristic past behavior of debt of various maturities represents a reasonable choice in view of his needs for earnings and liquidity; he then holds to this maturity structure irrespective of what is happening to interest rates or interest rate expectations. He plans regarding “capital value risk” on the basis of characteristic past fluctuations in prices of securities of different

maturity,⁹ but he also plans regarding yield on the basis of characteristic past yields on securities of different maturity.¹

What makes disturbing the refusal of some economists to recognize the possibility of nonspeculative investment behavior is not only that it so palpably exists, but that this whole matter has been threshed out at such length in practical discussions of investment strategy and "defensive" policies have been worked out quite explicitly *for the purpose of avoiding speculative decision-making*. The investor who does not want to, or is not permitted to, stand or fall on his ability to predict changes in interest rates (or stock prices) is not an oddity. Included are most financial institutions (though some of these will speculate within narrow limits) and many individual investors.

Defensive investment strategy clearly is conceived of as avoiding the use of specific expectations of the investor and being based upon the characteristic past performance of the assets involved. Its logic is that if there is not a major discontinuity between the experience of the future and that of the past, investment on this basis will produce average results that will be reasonably predictable, much more so than if they depended upon the judgment or forecasting ability of the particular investor. To ignore the actual basis on which such defensive investment is planned and read into it choices that were not made on the basis of expectations that do not exist represents a barren sort of cleverness.

9. This illustrates that the traditional concept of the liquidity of an asset as a quality estimated on the basis of its characteristic past performance rather than on the basis of specific expectations regarding future price behavior is a part of nonspeculative, or defensive, investment planning.

1. With regard to Wehrle's final point, that adverse liquidity conditions during the 1930's could have resulted from investors increasing "their estimates of the 'risk content' of all maturities" because of changed expectations (p. 606), I do not believe that it makes sense if interpreted in terms of a change in detailed expectations, for the reasons given above. The active speculator is not afraid of fluctuations and is little concerned with "liquidity."

However, the response of the nonspeculative investor to the disorderly debt markets of 1931 and 1933 undoubtedly was a factor contributing to liquidity deficiency in subsequent years. There would be different ways of putting the matter, depending upon how one defined the difficult concept of "liquidity." I would say that the experience with disorderly markets induced the defensive investor, operating on the basis of *past experience*, to lower his conception of the liquidity of debt of given maturity (including short-term, for that market also deteriorated). This investor would then shift in the direction of shorter maturities and more cash, if his preferred liquidity position and his estimated future cash needs were unchanged. Essentially the same point can be made in a commonplace way, by observing that investors who had been burned in the disorderly markets naturally sought a stronger defensive position. While the episodes of disorderly markets contributed to continued adverse liquidity conditions, it would seem a mistake to try to make of this the whole story of the unusually strong marginal demand for liquid debt during the 1930's.

We must admit that when we have recognized the two extreme types, the purely speculative or aggressive and the nonspeculative or defensive investor, we have only scratched the surface of the problem of the role of expectations in investment planning. There are, of course, many intermediate types of investors. Related to this, there are many different orders of expectations, from definite to vague, detailed to general, certain to tenuous. Undoubtedly there is some basic substratum of expectations, or what might be called expectations, underlying all economic decisions. There is, at least, an assumption that existing institutions will or will not continue in existence, and that future economic experience will be similar to or different from that of the past.

But these more basic foundations of economic decisions are probably better characterized as *planning assumptions* than as expectations. These planning assumptions are broad in character. For all but the most untrammelled and free-thinking investors, they rest largely upon the conventional interpretations of the lessons of past experience and operate through the various rules of thumb and canons of sound practice so important in most investment decisions. Planning assumptions can change, with far-reaching effects, because of new experience or new ways of thinking. An example is the shift now taking place in this country from a planning assumption of long-run price stability to one of unreversed inflation.

Exploration of the nature and variety of the planning assumptions underlying different types of economic decisions, their implications, and the shifts that have taken place in them could be a challenging task, one no less important, surely, than analyzing speculative behavior. In dealing with the general problem, as in considering its application to the interest rate structure, we must take the trouble to learn about actual decision-making processes and the thinking that underlies them. Armchair deductions of "logical implications" of decisions — those "implications" that produce ingenious and tidy abstract solutions — can teach us little.

THE EFFECT OF DEVALUATION ON THE PRICE LEVEL: COMMENT

By JAROSLAV VANEK

In his article in the May 1958 issue of this *Journal* Egon Sohmen demonstrates that under certain assumptions and conditions, using the Hicks-Kaldor-Scitovsky criteria, the "inflation argument" against devaluation and for exchange control may not be valid. In other words, it is a possibility that devaluation would lead to a general fall in domestic prices rather than to an increase. In the second part of his article, however, he argues that the conditions of the real world are such that in most cases devaluation would bring about a fall in domestic prices.

Sohmen's result is certainly a possibility, but in the real world it does not seem to be inevitable.

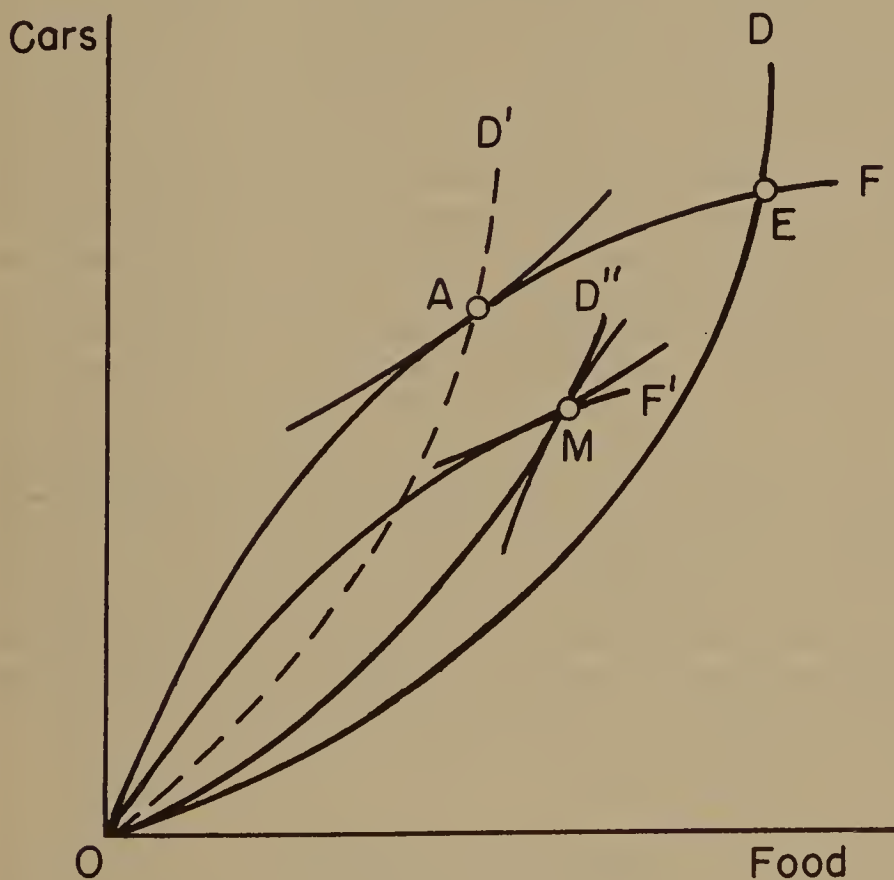
1. Although stated in a different manner the crux of the argument turns around the degree of optimum tariff. When the reciprocal demand curve of the restricting country is pushed by all possible restrictions, including quotas, tariffs, exchange control, etc., beyond the point of optimum tariff, then freeing of the exchange rate will be deflationary if the offer curve of the relaxing country (*OD*) settles at a point short of the "optimum" position; i.e., if *OD* before and after the change of policy intersects *OF* within *OA* in our diagram. If relaxation induces *OD* to move from *OA* into *AE*, the measure can be either inflationary or deflationary; and finally, if — with or without exchange control — *OD* intersects *OF* within *AE*, freely fluctuating rates will lead to an over-all increase of domestic prices.

Sohmen argues that the "deflationary" outcome is far more likely for two reasons: (a) the foreign reciprocal demand for most countries is highly elastic and consequently the "inflationary" stretch, *AE*, tends to be very small; (b) tariffs and quotas alone can usually be expected to go beyond the point of optimum welfare, and consequently relaxation of exchange controls would make the equilibrium point move only within *OA* — and this, as we know, is deflationary.

It seems that the first argument (a) is based on a consideration of the relative size of each particular economy in the world as a whole. Obviously, if all countries were producing only two homogeneous commodities and transportation costs could be ignored (as is usually assumed in simple international trade models), the foreign

offer curve would be almost perfectly elastic. But it is common knowledge that many countries (particularly the small ones, which Sohmen seems to have in mind) tend to specialize in a few out of the vast numbers of internationally traded products. Consequently it seems to be an oversimplification to assume that most countries face foreign offer curves which are highly elastic, at least in the short run.

As to the second contention (b) I first would like to make a minor point. Tariffs, quotas and controls are not additive, as Sohmen contends. Controls and quotas, for example, are nonadditive, and



whichever is higher will determine the equilibrium. A simple parallel may illustrate this case: A 10 per cent controlled price reduction which would lead to a 10 per cent fall in supply and rationing which would lower physical consumption by 10 per cent will result in only a 10 per cent fall of sales rather than in one of 20 per cent. Second, and more important, Sohmen seems to operate with an intuitive notion of the "optimum degree of tariff" when he asserts that tariffs alone have carried each country within OA in our diagram. To my

knowledge, no one has ever even tried to estimate the optimum point. Statistical and conceptual difficulties are such that to do this is virtually impossible. It is difficult to conceive of the "detailed empirical investigation" which would, according to Sohmen, settle this question decisively.

2. In my second comment I would like to follow further Sohmen's purely deductive reasoning. It may well be that even without any precise, explicitly stated functions, governments of different countries actually play a "semi-rational" game, trying to reap from their trading partners as much as they can by a process of retaliation. It is a commonplace that in such a case equilibrium will be possible at a point, such as *M* in the diagram, at which each country's offer curve is tangent to the other country's social indifference curve. It seems that if there is any correspondence between our abstract models and the real world, then it is this situation which corresponds most closely to reality. If this is the case, however, it is immediately apparent that unilateral relaxation of any restriction (freeing of exchange rates in our case) will necessarily lead to inflation in the relevant country.

3. Pursuing his argument, Sohmen contends that "The fall in the price level . . . must be all the more pronounced *the larger is the extent of depreciation* necessary to establish equilibrium in the exchange market . . ." This is true *only* if the depreciating country's offer curve intersects the foreign offer curve within *OA* both before and after depreciation — a case which Sohmen seems to take for granted. However, if no specification is made as to the position of equilibria before and after depreciation, no simple relation can be established between the extent of depreciation and the change in domestic prices. In this case, possibly an *important* reduction of the exchange rate could be inflationary while a *small* change of the rate in the same direction would reduce domestic prices.

FURTHER COMMENT

By RANDALL HINSHAW

Mr. Sohmen has written an interesting theoretical article¹ on a subject which he recognizes is of much more than purely theoretical interest. The effect of devaluation on the price level (in particular, on the cost of living) is a matter of profound concern to officials

1. Egon Sohmen, "The Effect of Devaluation on the Price Level," this *Journal*, LXXIII (May 1958).

contemplating such action, and the views that economists express on this question are likely to receive more serious attention in official circles than is usually accorded to academic opinion. It is therefore important that such views be as carefully considered and as plainly stated as possible.

At the outset, let me say that I am completely in sympathy with Sohmen's belief in the need for realistic rates of exchange, both as a matter of enlightened national self-interest and as a prerequisite for a satisfactory international payments system. Moreover, I fully share his interest in exposing any weaknesses in accepted theorizing that may deter officials from proceeding with devaluation in cases where such action may be badly needed. The opposition to devaluation from nonacademic sources — notably from high-cost producers sheltered by import restrictions imposed as an alternative to devaluation — is hard enough for officials to deal with without having also to cope with bureaucratic opposition based on impressive but dubious arguments issuing from academic circles. But in disassociating himself from some widely held views, Sohmen has staked out a new position which, because it may lead to misdirected hopes in official quarters, seems to me in need of reorientation in certain important respects.

Sohmen rightly points out that, where a measure of international balance has already been achieved by means of import restrictions, devaluation need not lead to a rise in the *retail* price level of imported goods, and may indeed lead to a fall. If devaluation is followed by a healthy increase in earnings of foreign exchange, it should be possible to reduce or remove the import restrictions, thus permitting an increase in the volume of imports. Under such conditions, the retail import price level would tend to fall.²

Up to this point, Sohmen is on firm ground. From the reasoning just outlined, however, he draws two conclusions which I find more difficult to accept. First, because of his demonstration that retail import prices may fall in cases where devaluation is substituted for import restrictions, he concludes that the general domestic retail price level in such cases is also likely to fall. Second, Sohmen apparently regards any tendency of the general price level to rise following

2. This is the same proposition I made in 1951 in an article published in this *Journal* ("Currency Appreciation as an Anti-inflationary Device," this *Journal*, Nov. 1951). On the basis of the foregoing reasoning, I argued that, just as devaluation may lead in certain cases to a fall in the retail import price level, so may currency appreciation lead to a *rise* in retail import prices. As Sohmen indicates, the same point has been made in print by Ragnar Nurkse and by E. M. Bernstein.

devaluation not only as unlikely (assuming appropriate monetary and fiscal policies) but also as undesirable, and in fact contends that where such a rise occurs the "potential welfare" of the devaluing country is reduced. Indeed, his case for devaluation seems to fall to the ground unless it can be shown that devaluation tends to result in a decline in the general price level.

These conclusions, if I have understood them correctly, seem to me seriously in need of modification. In the first place, for reasons indicated below, I would maintain that, even in cases where devaluation is followed by a fall in the retail import price level, the general retail price level in the short run is likely to rise or at best to remain constant.³ In the second place, instead of regarding a rise in the general price level as an unfortunate or undesirable by-product of devaluation, it seems to me abundantly clear that in most cases a price-level rise is an important if not essential part of the mechanism of international adjustment and that attempts to frustrate such a price rise (other than by *deflationary* monetary or fiscal policies) are likely also to frustrate the return to international balance.⁴

The propositions which I have just set forth will perhaps be accepted by a substantial number of the profession without further demonstration. Since, however, they differ from the views of Sohmen, and since the second proposition differs from the widely held view that any price increases following devaluation tend to reduce or remove the competitive benefits conferred by such action, I shall endeavor to justify my position as lucidly as I can. To this end, I shall first submit a few generalizations as background. These generalizations are not offered as anything new — most of the points, I am well aware, have been made more elegantly by others — but are presented as essential parts of the argument.

1. International disequilibrium is generally characterized by, and is in part the result of, an inappropriate relationship between a country's level of money incomes and the country's general retail price level.⁵ As here used, the term "general retail price level" (hereafter shortened to "price level") embraces both the retail price level of consumption goods and services (i.e., the cost of living) and the

3. I agree with Sohmen that, in the long run, devaluation may lead to a decline in the price level by promoting more efficient investment.

4. It should be needless to state that I am referring here to price-level increases which are a direct result of devaluation and not to those resulting from a continuation or resumption of inflationary policies.

5. It is also generally characterized by, and is in part the result of, an inappropriate relationship between the price level of international goods (imported goods and exportables) and the price level of home goods, but I prefer at this stage to concentrate on the first relationship, since it is so often overlooked.

retail price level of investment goods. It is affected (in most countries strongly affected) by what happens to the retail price level of *international* goods — that is to say, both imported goods and exportable goods, since a substantial proportion of exportable goods is normally consumed at home.

2. Where the international disequilibrium takes the form of a chronic deficit, the price level, as thus defined, may be regarded as too low for the level of money incomes. Real income (i.e., money income deflated by the price level) is therefore too high, leading to excessive expenditure both on imports and on exportable goods.⁶ The result is a deficit in the international accounts. In the words of the postwar cliché, the country is “living beyond its [current] means” — a situation made possible either by living off past savings (i.e., by losing monetary reserves) or by obtaining special assistance from abroad in the form of grants or loans.⁷

3. In situations of this kind, as Sohmen reminds us, a deficit may be reduced or removed by means of import restrictions. Whether these take the form of tariffs or of quantitative limitations, the effect is to raise the retail price level of imports (and of import substitutes) and thus to raise the retail price level as a whole. This action, if carried far enough, restores the price level to an appropriate relationship to the level of money incomes (and thereby reduces real income and expenditure to an appropriate level), but it is an inferior solution since, as Sohmen points out, it represents a move in the direction of autarky and therefore in the direction of lower productivity.

4. A better prescription for getting rid of a deficit, most would agree, is devaluation. Where balance has already been achieved by means of import restrictions, devaluation can alternatively be employed to get rid of the restrictions. In addition, there are frequently cases where a deficit is reduced but not removed by import controls. Here devaluation may be used merely to get rid of the residual deficit or may be more ambitiously employed to remove both the deficit and the restrictions. To avoid confusion, each of these cases should be carefully distinguished.

It may be helpful in this connection to employ an illustration

6. Students of the late Professor Frank D. Graham will recognize my debt to him in this analysis, which also owes much to many others.

7. To be *chronic*, this type of disequilibrium requires either (a) a persistent inflationary process (in cases where the deficit is financed by an outflow of monetary reserves) or (b) continuous financial assistance from abroad. In the former case, the deficit can last as long as the country is able and willing to lose reserves; in the latter case, the deficit can last as long as financial assistance from the outside is forthcoming.

that I have found useful in the classroom. It is an illustration designed to isolate the essential elements in the international adjustment mechanism, and can be modified to introduce as many complexities as are desired.

Let us imagine a very small country in the form of an island located not far from a populous mainland. To make matters as simple as possible, let us suppose that the island is simply an enormous rock on which a few hundred people have lived for many generations. Let us assume that the surrounding waters abound in fish and that the island has a fresh water lake, so that subsistence (fish and water) is barely possible in the absence of international trade. We may assume, however, that a thriving trade with the mainland has long been carried on, in which exports of fish have been exchanged for a wide variety of imports that make the difference between an agreeable life and a bare existence for the islanders.

In addition to fishing, which we shall assume is carried on by wage earners working at a standard wage rate, we may suppose that there are also some purely domestic industries, which for simplicity we shall assume are service industries, such as laundering. Such work, we shall assume, is also done by wage earners working at the standard wage rate.

As a preliminary assumption, which can be modified later, we shall suppose that the population of the island is so small in relation to the population of the mainland that variations in neither the exports nor the imports of the island have a perceptible effect on mainland prices. Finally, let us assume that the island has an independent fiscal and monetary system. To make matters concrete, let us call the currency unit of the mainland the dollar and the currency unit of the island the peso. We shall suppose that, over the years, the island has acquired a substantial monetary reserve in the form of dollars.

The price level of the island in this situation will reflect the prices of international goods — both imports (e.g., clothing) and exportables (fish) — and the prices of domestic output (e.g., laundry). Under conditions of international equilibrium, (a) the relationship between the price level and the level of money incomes, and (b) the relationship between the price level of international goods and the price level of domestic output will be such that the island's dollar receipts will be equal to the island's dollar expenditures.

But a disequilibrium in this situation is easily imaginable. Let us suppose that, during a period of internal and external equilibrium, the island decides to inaugurate a generous system of old-age pensions and to finance this by printing money. Money incomes per period

are thereby increased, leading the islanders to increase their expenditures both on international goods and on domestic output. The result of the increased domestic expenditure on fish is a decline in exports, and this, together with the increase in imports, produces a deficit in the balance of payments. Since, under our assumptions, the price level of international goods is not significantly affected by the inflationary process whereas the price level of domestic output is free to rise, a transfer of factors from the fishing industry to domestic industry is encouraged, thereby producing a "structural" disequilibrium. The malady can continue as long as money incomes remain inflated because of inflationary policies and as long as the island is able and willing to finance the external deficit.

In this situation, two things are wrong: (1) the price level is too low in relation to the level of money incomes (i.e., real income is too high), and (2) the price level of international goods is too low in relation to the price level of domestic output. With regard to the first point, real expenditure is too high, and takes the form of excessive expenditure on imports and on exportables. With regard to the second point, too much of the island's production is in the form of nonexportables (e.g., laundry) and not enough in the form of exportables (fish).

An obvious prescription — and an essential part of any permanent solution — is to stop the inflation. If nothing else is done, however, this means letting money incomes fall to their equilibrium level since, once the inflationary process is halted, the money-income leakage via the balance of payments is no longer offset by the money-income creation via the printing press. Thus the deficit automatically tends to disappear. But if the disequilibrium has been of long standing, such a policy may be unduly painful, possibly entailing a considerable amount of unemployment.

An alternative to a lower level of money incomes (but *not* an alternative to stopping the inflation) is devaluation. This corrects the disequilibrium *by raising the price level*. Devaluation raises the price level of international goods as a group — both imports and exportables. Indeed, in our illustration, it raises the international price level (in pesos) by about as much as the change in the exchange rate (i.e., by about as much as the rise in the price of dollars in terms of pesos), since the dollar prices of international goods, under our assumptions, are not appreciably affected by changes in demand or supply on the island.⁸

8. The rise in prices clearly applies to exportables as well as to imported goods; if the price of fish on the island were not bid up by an amount approxi-

The rise in the international price level has two effects. In the first place, by raising the average price level, it lowers the island's level of real income and expenditure, the reduction in expenditure being concentrated on international goods.⁹ Mainlanders, in effect, are permitted to outbid the islanders for an increased quantity of the island's output of fish, while islanders reduce their purchases (in terms of quantity) both of fish and of imported goods. Since dollar prices are not significantly affected by the devaluation, the island's dollar earnings clearly rise and its dollar expenditures clearly fall. In the second place, by raising the price level of exportables in relation to nonexportables, devaluation makes fishing operations more profitable than laundering (assuming that the rate of earnings in the two industries had previously been the same). Consequently, not only will more fish be exported out of existing fish production, but fishing operations will be encouraged to expand. To attract workers from domestic industry, the fishing industry may have to offer somewhat higher wage rates, but provided the level of money incomes is prevented from rising significantly — or, at any rate, from rising by as much as the rise in the price level — devaluation will promote a return to equilibrium both by reducing real expenditure on international goods and by redirecting production in favor of exportables.

In the situation we have just considered, devaluation clearly involves a rise in the price level and — if money incomes are prevented from rising proportionately — a fall in real income.¹ In the latter context, it may also be said to involve, in Sohmen's language, a fall in "potential welfare," though only in the sense that the island enjoys a somewhat lower standard of living because it is no longer living beyond its current means. Whether real income will have to fall slightly or substantially will depend on the magnitude of the deficit in relation to the level of income. If this ratio is low, as is generally the case even in situations of serious disequilibrium, the

mately equal to the rise in the price of dollars in terms of pesos, island fish producers would sell all their output on the mainland.

9. This part of the adjustment process will, of course, be frustrated if the level of money incomes is permitted to rise by as much as the rise in the price level. It should be the firm aim of the authorities to prevent this — if necessary, by supplementary fiscal and monetary measures.

1. It is sometimes argued (for example, by Joan Robinson) that income effects can be ignored in the adjustment process, since they emerge as a *result* of a change in the balance of payments, and thus can affect only the magnitude, but not the direction, of the change. This is true of certain income effects, but it should be emphasized that the real income effect here referred to will occur (subject to the proviso already indicated) even if there is no immediate change in the balance of payments following devaluation. Thus this income effect can play a powerful role in the adjustment mechanism.

required reduction in real income will be moderate. In any case, it should be noted that, in the case under consideration, the adjustment process does not involve a "secondary burden" in the form of a deterioration in the commodity terms of trade. That is to say, the ratio of export prices to import prices remains the same after devaluation as before, since both price levels (in pesos) rise by the same proportion.²

Sohmen may properly object at this point that the case which I have thus far outlined does not meet his argument, which is concerned mainly with devaluation as a substitute for import restrictions previously in force. Let us therefore modify our illustration to take this into account. Let us suppose that, instead of originally dealing with the deficit by devaluation, as assumed above, the island first resorted to import restrictions and was thereby successful in completely removing the deficit. As previously indicated, the effect of such restrictions, whether tariffs or quotas, is to raise the retail price level of imports and, as a result, to restore a relationship between the general price level and the level of money incomes which can be maintained without an external deficit (and which reflects the necessary reduction in real income). Unlike devaluation in our first case, this solution involves a very real "secondary burden," not in the form of a deterioration in the terms of trade, but in the form of a qualitative deterioration in living standards resulting from the shift toward autarky. In the case under consideration, the island is forced — quite unnecessarily — in the direction of living solely on fish and water. This is the "austerity approach" for achieving international balance, and I agree entirely with Sohmen in questioning its wisdom under most conditions.

Let us suppose, however, that external balance has been maintained for some time by means of import taxes or quotas and that the island has finally decided, as an alternative approach, to devalue. What will happen to the price level in this case?

First of all, it would seem clear that no cut in real income is indi-

2. In dollars, both price levels remain (virtually) unchanged. In the case of a larger country, as I have tried to show elsewhere (*op. cit.*, pp. 456-58), devaluation may be followed by a change in the terms of trade (either an improvement or a deterioration) if the effect of the devaluation on the foreign prices of imported goods (a downward effect) is greater or less than the effect of the devaluation on the foreign prices of exportables (also a downward effect). In any case, it should be clear from the foregoing analysis that a change in the terms of trade is not a necessary part of the adjustment process; on the contrary, a deterioration, if it occurs, simply makes the process more difficult by adding a secondary burden.

cated in this situation, since the necessary reduction has already been achieved by means of import restrictions.³ Consequently, there would appear to be no need for a rise in the price level as part of the adjustment process (the rise having already occurred as a result of the restrictions on imports). At the same time, I find it difficult to agree with Sohmen that there would be a tendency in the short run for the price level to fall.⁴ If I am not mistaken, the correct conclusion in this type of situation is that devaluation, while affecting some retail prices downward and some retail prices upward, would not reveal a consistent tendency either to raise or to lower the average retail price level.

Assuming that balance has already been attained by import restrictions, and assuming that devaluation is followed by a substantial increase in earnings of foreign exchange, it would be possible to reduce or remove the import restrictions and thereby to permit a substantial increase in the volume of imports. Sohmen and I agree that this would tend to lower the retail price level of imports. At the same time, however, the price level of exportables, in domestic currency, would rise as a result of devaluation. In our illustration, a fall in the retail price level of imported goods (e.g., clothing) would be offset by a rise in the retail price of fish — an important item in the island's cost of living as well as an export item.⁵ Under conditions of external balance both before and after devaluation, the presumption is surely that these two influences on the average price level would tend to cancel out.⁶ Indeed, where balance has not been completely achieved by import restrictions, devaluation would be expected to involve some rise in the price level as part of the adjustment process required for eliminating the residual deficit. Since devaluation is almost invariably resorted to by countries confronted both

3. I am overlooking here the possibility of a deterioration in the terms of trade as a result of removing the import restrictions, since I agree with Sohmen that the degree to which a country can influence its terms of trade in this fashion is in most cases very limited. In the case of very small countries, any such effect would be microscopic.

4. Of course, if the import restrictions were removed *without devaluing*, the price level would tend to fall, but at the cost of reopening the deficit.

5. The price level of domestic output (e.g., laundry) would presumably remain unchanged.

6. I do not wish to be dogmatic on this point; as Sohmen shows, the results are affected by the type of index number used. Moreover, from the standpoint of policy, interest may be centered on the cost of living rather than on the over-all price level. If a country imports mainly consumption goods and exports mainly investment goods, the cost of living may fall following devaluation (assuming balance has already been achieved by restrictions); on the other hand, if a country exports mainly items which are prominent in its own pattern of consumption while importing mainly investment goods, the cost of living will tend to rise.

with a persistent deficit and with a legacy of import restrictions imposed to reduce the deficit, a rise in the price level would normally be expected where devaluation is introduced either with the modest purpose of getting rid of the residual deficit or with the more ambitious objective of eliminating both the deficit and the restrictions.

While I doubt that there is an important category of cases where officials can count on a fall in the price level as a (short-run) result of devaluation, I am in full agreement with Sohmen that, where devaluation is substituted for tight import restrictions, the country taking such action can look forward to a *qualitative* improvement in real income. Devaluation as a substitute for import restrictions represents a move away from autarky — in our illustration, a move away from a fish-and-water existence. Thus, in such cases, devaluation may be abundantly justified even if real income (as measured by the relation of money income to the price level) shows no increase or even shows a decline.

I would like to end this discussion on another note of agreement with Sohmen. In the background of his argument is the implicit assumption that devaluation, when not frustrated by inappropriate policies, can be expected to result in a substantial increase in foreign-exchange earnings. Despite the many doubts expressed in recent years, this seems to me a reasonable assumption in the great majority of cases. The point will doubtless be conceded by all in the case of very small countries. But smallness is a matter of degree, and most countries account for only a small percentage of world demand and world supply. Moreover, small countries in external difficulties seem to resist devaluation as frequently as large countries. The reason they almost invariably cite for such resistance — particularly if their exports are mainly primary commodities — is that world demand for their products is inelastic and that devaluation would therefore result in a decline in earnings of foreign exchange. This argument, which has impressed many a bureaucrat, usually confuses (a) world demand for world output of the commodities in question with (b) world demand for an individual country's output of such products. Even if (a) is highly inelastic, as is often the case, (b) may prove highly elastic, particularly if devaluation is undertaken unilaterally and is not followed by retaliation. Without reopening the question of their reliability, it may be pointed out that most statistical investigations of the price elasticity of demand for foreign output shed no light on this matter, since they are concerned with output from the rest of the world as a whole.

REPLY

By EGON SOHMEN

I warmly welcome the comments of Messrs. Hinshaw and Vanek; they should serve further to elucidate what I was trying to say in perhaps too concise language. There is virtual unanimity on theoretical matters. As far as Professor Hinshaw's comment is concerned, I would be hard put even to discover a significant difference of opinion on questions of empirical fact.

Let me restate once more one slightly uncommon feature of the approach I had chosen; it might otherwise give rise to misunderstanding.

It seems to me absolutely essential for clarity to consider separately (a) the effect of devaluation per se (with the balance on current account unchanged), and (b) the consequences of a change in the balance on current account (with the exchange rate held constant). The balance on current account is understood to be measured in terms of foreign currency wherever ambiguity arises.

For case (a), the basic principle of comparative advantage guarantees that the index number inequalities can hold if all the auxiliary conditions specified in my paper are fulfilled.¹ It follows that a certain (although, of course, limited) improvement in the trade balance (case (b)) may be possible after devaluation while the index numbers still indicate a fall in the price level.

All this involves comparison of different positions of static equilibrium. Hinshaw is quite correct in pointing out that events in the short run, before all adjustments have taken place, may tell a different story. But more about this later.

For policy decisions on exchange rate adjustment, I expressed the view that effects on the price level can be legitimately compared only under assumption (a) that the trade balance does not change.² For many people, devaluation without a change in the balance of trade may appear like Hamlet without the Prince. It may be argued that it was really a substantial improvement in the trade balance after devaluation which economists have implicitly regarded as the cause of its inflationary effect.

But in opposing devaluation for this reason, they seem to have in mind a make-believe world in which governments have a free choice in determining the size of their trade deficits. As a picture of

1. *Op. cit.*, pp. 275-76.

2. *Op. cit.*, p. 282.

reality, the assumption of an unchanged balance on current account does not seem to me to be far-fetched at all. The foreign-exchange reserves of precisely those countries who would need them most are severely limited indeed; and capital inflow, alas, does not come forth at the wave of a wand. With or without devaluation, these restraints enforce a certain maximum trade deficit per unit of time which simply cannot be exceeded for very long. If inertia, lobbying, or superstition prevents the exchange rate adjustment required to hold the deficit down to the feasible level, brute force in the form of direct controls will have to be applied. Everybody who has witnessed exchange overvaluation in action will easily recognize this portrait. It is on this situation that my case was built.

Advertising overvaluation as an irreparable, deep-seated disequilibrium may be a highly effective trump card in the hands of a government looking for financial assistance from abroad. An additional capital inflow may be secured in this way to permit an increase in the feasible trade deficit. I did not fail to pay tribute to this relatively recent and ingenious addition to the arsenal of beggar-thy-neighbor policies.³ What makes it all the more effective is that it is usually applied in perfectly good faith.

But there is another important factor, apparently widely neglected, that works in the opposite direction: overvaluation is one of the most powerful deterrents to private capital movements. Exchange controls can to a certain extent prevent massive capital flight; the kind of controls that could induce large-scale capital *inflow*, on the other hand, has yet to be invented. A realistic appraisal would, I believe, lead to the conclusion that governmental grants are frequently quite insufficient to make up for the private capital flows repelled by overvaluation. Where this is true, devaluation to the equilibrium level will even permit an *increase* in the feasible trade deficit per unit of time.

Let us return to the question of short-run versus long-run consequences. Assuming that the balance on current account remains constant, the index-number criteria are liable to conflict as long as no adjustment in production takes place.

Figure I illustrates a possible outcome when trade is exactly balanced before and after devaluation. Before devaluation, production is indicated by P , consumption by C_1 , the internal equilibrium price ratio by p_1 . It differs from the external terms of trade t_1 because foreign trade is assumed to be limited by restrictions.

3. *Op. cit.*, p. 282. Again, I want to emphasize my sympathy for international co-operation wherever needed. But I would prefer to see the right thing done for the right reason.

If we assume that all controls are removed after devaluation, the internal and external price ratios must coincide ($p_2 = t_2$). Production still takes place at P as long as resources are not reallocated. Let us assume that trade moves consumption to C_2 . When national income (in money terms) is unchanged, the Laspeyres index shows a rise, the Paasche index a fall in the price level.⁴

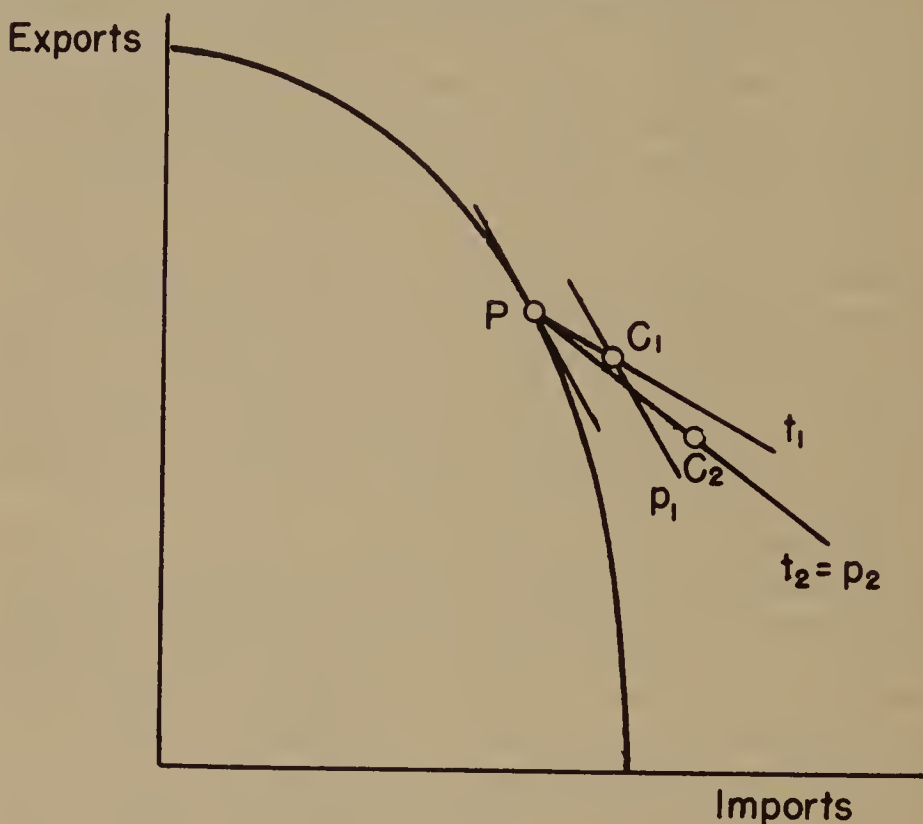


FIGURE I

Time series of cost-of-living indices are invariably of the Laspeyres variety, owing to the formidable work involved in establishing base weights. Since factors of production show immobility in the short run, it is indeed quite likely that the official (Laspeyres') chronicle will initially record a price rise in most cases; provided, of course, that the price index of the sample commodities moves approximately in proportion to the aggregate index.⁵ But this is only due to the accident that Paasche indices are excluded for short-run comparisons because of the practical difficulties they present. Even for moderate

4. The Laspeyres index remains constant for a country that cannot affect its terms of trade.

5. See the remarks on pp. 276-77 of my previous paper.

demand elasticities, they would just as certainly indicate a fall in the price level.

The commonsense explanation for the behavior of price indices is simple. In the market basket bought after devaluation, commodities whose prices have fallen will be favored, by and large. These cheaper goods therefore enter into the Paasche index with greater weights; in contrast, the Laspeyres index places more weight on those that have become relatively more expensive.

Toward the end of his comment, Hinshaw stresses the significance of "qualitative" improvements in real income. The term is entirely justified if we operate with Laspeyres indices. When the Paasche index is used, the improvement is fully measurable and quantitative, for the reasons explained above. This outcome is strongly reinforced if certain commodities only become available after devaluation. It might even be argued that their predevaluation price would have to be set at infinity.⁶

The moral to be drawn from the fact that devaluation possibly results in a fall of the price level only after a certain time lag will probably depend on a man's temperament. Experience seems to show that many economists subscribe to the conservative view that any action which combines benefits in the long run with initial inconveniences should be avoided. Another school, and one towards which I feel considerably more sympathetic, would suggest that there is all the more reason to take those measures early whose fruits are known to take some time to mature. As far as I can judge, Hinshaw would probably join me in this view, although I am afraid we may belong to a minority.

Except for two minor details, there is no difference of opinion between Mr. Vanek and myself on points of theory. The first of these concerns the additivity of different forms of restrictions in foreign trade; my claims in this respect (p. 279) were stated uncautiously, to say the least, and are quite willingly withdrawn. They do not affect the rest of the argument anyway.

Secondly, I do not favor his restatement of my main argument in terms of community indifference curves. Where it is unnecessary or unwarranted to assume the feasibility of interpersonal welfare judgments, their use ought to be excluded for all analysis exceeding minimal requirements of rigor. All they can legitimately show can be demonstrated in terms of individuals' indifference maps, while it is

6. A vivid example due to Professor Haberler is the substitution of a Mediterranean vacation, inaccessible to a resident of austere postwar England, for a summer in a rainy British resort.

all too easily forgotten what they *cannot* show. It may be, however, that his presentation is helpful for readers who are not too familiar with index number reasoning. Inevitably, all the subtle ambiguities of index number comparisons are lost, of course.

The quotation in the last paragraph of Vanek's comment is taken from a context in which it is assumed that the regular trade restrictions (i.e., those that remain even after the devaluation) already exceed the "optimum" level. I fully agree that the statement would be nonsense without this qualification.

The question whether, under our common assumptions, a fall or a rise in the price level is more likely in any given situation can, by its very nature, not be decided by theoretical reasoning alone. On the other hand, what kind of empirical evidence should be considered?

I share Vanek's view that all that attempts at econometric measurement of elasticities of internationally traded commodities have so far demonstrated is their hopelessness, at least with the tools presently available. A much simpler as well as potentially more promising method is, I believe, an indirect appraisal by what might be christened "revealed reference" from the actual history of recent devaluations.

The attitudes of many economists toward other issues throw a dubious light on the strength of their professed elasticity pessimism. While many of our British colleagues, e.g., have repeatedly discounted devaluation of the pound as an effective measure to relieve pressure on sterling, all fears of low elasticities seemed to be forgotten in the campaign of 1957 to have Germany appreciate the mark. Sufficient faith in their previous convictions would have compelled them to wonder whether devaluation of the German mark might not be a more appropriate step to reduce the German export surplus.

Not infrequently, elasticity pessimists are strong believers in trade restrictions for the protection of infant industries. Whenever foreign reciprocal demand is even slightly inelastic, however, there is a presumption that the physical volume of imports will increase after tariffs are raised.

Oddly enough, (or perhaps not so oddly?) it appears exceedingly difficult to find a clear-cut example in support of the orthodox belief. The world is replete, of course, with cases in which devaluation was undertaken under the pressure of inflation — creeping, trotting, or galloping — and where the inflationary trend continued merrily at about the same rate after devaluation. But one can hardly blame the latter for this result under such circumstances. Devaluation has, furthermore, resulted in an improved balance on current account in

TABLE I

	Price Index		Wage Index April 1945 = 100	Employment Index (seasonally adjusted) 1948 = 100	Discount Rate (per cent)
	Wholesale 1938 = 100	Retail 1938 = 100			
Year's Averages					
1952	822	669	723	99.7	
1953	776	664	723	98.6	
1954	811	682	759	102.0	
Monthly Figures					
1953					
Jan.	790	670	723	98.1	5
Feb.	786	670	723	98.1	1
March	767	665	723	98.1	6
April	750	660	723	98.2	
May	760	654	723	98.3	
June	788	659	723	98.5	
July	777	658	723	98.6	
Aug.	791	661	723	98.6	
Sept.	773	665	723	98.7	4
Oct.	775	667	723	99.8	
Nov.	778	667	723	99.0	
Dec.	778	667	723	99.2	
1954					
Jan.	794	672	740	99.5	
Feb.	781	673	740	99.9	
March	784	669	740	100.2	
April	788	672	751	100.6	
May	794	675	763	101.0	
June	800	684	763	101.6	3.5
July	844	686	764	102.1	
Aug.	825	684	765	102.8	
Sept.	822	686	764	103.3	
Oct.	825	687	770	103.8	
Nov.	838	690	777	104.3	
Dec.	839	691	778	104.8	3.5

Source: *Monatsberichte des Österreichischen Instituts für Wirtschaftsforschung*, Vienna, Vol. 26, 27 and 28 (1953, 1954, 1955), and the Statistical Supplements to the *Monatsberichte* for the same years.

most instances, whereas a realistic comparison would, as emphasized before, require that the latter remain unchanged between the two time periods under investigation.

I should like to present the case history of a recent devaluation in which at least the first requirement (of relative price stability before the measure) was fulfilled. The reader is invited to draw his own conclusions.

In May 1953 Austria established a single rate of exchange of

Schillings 26. — per U.S. \$ in lieu of the two previous rates of 21.36 (for most commercial transactions) and 26. — per U.S. \$ (for tourist trade and other invisibles and certain luxury items).

The balance on current account improved dramatically. For the first time in the history of the Austrian Republic (i.e., since 1919), the value of commodity exports roughly equaled the value of commodity imports in 1953.⁷ Gold and exchange reserves increased by 180 million dollars during 1953 and another 110 million during the first nine months of 1954, while American aid dropped from 105 million dollars in 1952 to 40 million in 1953. A few months after the devaluation, the discontinuation of all further aid was announced — perhaps a persuasive example of the Macchiavellian merits of currency overvaluation.

Austria's terms of trade deteriorated by 13 per cent between 1952 and 1953 according to the estimates of the Austrian Institute of Economic Research. This development was, however, largely independent of the devaluation.

As shown by the figures below, the price as well as the employment level remained virtually unchanged for a whole year in spite of the elimination of the import surplus and in the face of a strongly *expansionary* monetary policy: the central bank discount rate was lowered from 6 per cent to 4 per cent in two successive steps during 1953, and the money supply increased by about 20 per cent.

YALE UNIVERSITY

7. Both exports and imports were about 17 per cent of the Gross National Product in 1953.

A NOTE ON PROFESSOR SOLOW'S GROWTH MODEL

By JOHN BUTTRICK

In the February 1956 issue of this *Journal*, Robert M. Solow presented an ingeniously simple yet extremely useful model for the examination of various aspects of the problem of economic growth.¹ It does not seem to be generally realized, however, that this model, after a minor change, may be used to make sense out of the assertions that: (a) a large initial dose of investment may be required before an underdeveloped country can overcome existing barriers to growth; (b) it may pay to adopt capital-intensive production methods even if they appear to be "noneconomic"; (c) in contrast to (a) above, something very like spontaneous combustion sometimes occurs, i.e., for no apparent reason the growth process becomes self-perpetuating. This feature of the model may be used as an important element in an explanation of "the industrial revolution."

To proceed with the argument, let me first reproduce Solow's equation (6):

$$\dot{r} = sF(r,1) - nr$$

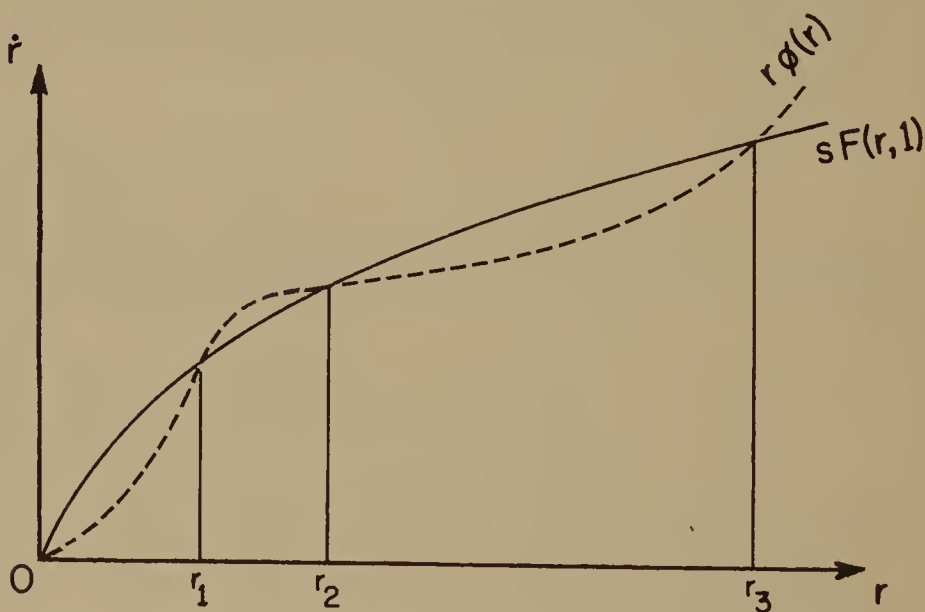
where, it will be remembered, r is the ratio of capital to labor, i.e., $r = K/L$; \dot{r} is the time derivative of this ratio; s is the average and marginal propensity to save; n is the growth rate of the labor force; and $Y = F(K,L)$ is the production function (assumed homogeneous). In the above equation, this function is written in terms of real output per worker, i.e., $Y/L = F(r,1)$. Second, let me treat n as a function of r . This makes sense because the capital-labor ratio is a reasonably good indicator of the country's level of development. It appears, for example, that the rate of population growth first increases as agriculture is replaced by industry, slows down as the growth of industry brings urbanization, and possibly increases once again when development has proceeded far enough to bring affluence in its wake.

Alternatively n may be treated as a function of Y/L as Solow does in his Figure IX. (There is, obviously, a direct relation between r and Y/L which is suppressed in the equation given above.) As income per worker rises with development, the growth rate of the (fully employed) labor force first rises and then diminishes. When really

1. "A Contribution to the Theory of Growth."

high income levels are reached, however, as occurred in the United States after World War II, the labor force once again begins to grow rapidly. To rationalize this supposition one could argue, à la Malthus, that an increase in income, with occupation or social class held constant, will result in a population upsurge. On the other hand, a change in the composition of the population such that the proportion in the more skilled occupations or higher social classes is increased will, with income constant, result in a decrease in the birth rate. During the process of development, it appears likely that this second sort of change dominates the first for a long time. As change in the social class structure slows down, however, continued development will cause further increases in income so population will once again begin to rise.

Either of these suppositions permits the construction of the diagram below in which the solid line represents $sF(r,1)$ and is reproduced from Figure I in Solow's article. The dotted line represents $r\phi(r)$, with $n = \phi(r)$, which is drawn to conform with the arguments just given. Notice that r_1 and r_3 are capital-labor ratios that possess stability properties while r_2 is unstable.



To interpret the process of development or industrialization with the aid of this diagram, suppose an economy is initially in the "agricultural stage" with $r < r_1$. Growth will occur until r_1 is reached but then will cease in the sense that a stable equilibrium capital-labor ratio has been reached and real output per worker will be constant over time. Attempts to increase output per worker will fail unless

$r > r_2$ can be achieved, i.e., small increases in r will only temporarily push r above r_1 . As the figure is drawn, a massive change in r is required to get over the hurdle to development denoted by $(r_2 - r_1)$.² Once r_2 is passed, however, "automatic growth" will occur until r_3 is reached.

With this picture of a development process before us, we can see immediately three general ways by which the barrier $(r_2 - r_1)$ may be overcome: (1) the $r\phi(r)$ function could be lowered; (2) the $sF(r,1)$ function could be raised; (3) r could be increased suddenly by an exogenous injection of additional capital, e.g., a foreign loan or gift or perhaps the discovery of new resources. Either of the first two changes would, in the course of time, push the solid curve up or the dotted curve down until the r_1 and r_2 intersections vanished and only r_3 remained. Should this happen, growth would become automatic until r_3 was reached.

Perhaps the most intriguing of these possibilities is (2). The appearance of the Protestant Ethic, for instance, would increase s and raise the solid curve. Technological change, even of a "neutral" sort, would be equivalent in effect. We can imagine these sorts of changes gradually raising the solid curve until finally the camel's back is broken, i.e., the two curves become tangent. Then, suddenly and "for no apparent reason," the economy would find itself starting on an automatic growth path toward r_3 . This interpretation of the Solow model, it should be noticed, provides the beginnings of a theory of "the industrial revolution" that does not require any special set of fortuitous and exogenous events. This is a real advantage!

A government can, in this model, itself introduce changes equivalent, in effect, to those just discussed. By taxing or inflating the currency, the government may be able to hire resources away from noninvestment uses and itself increase the capital stock. In some cases, borrowing from the public might work to the same end. The Russian argument that investment in heavy industry may be an effective way of overcoming the hurdle $(r_2 - r_1)$ is obvious. In principle, population control or massive foreign loans (or gifts) would also work. As an economist, however, I am unable to say much about these possibilities. Nonetheless, notice that it may be possible to partition an economy (or the world) into separate compartments according to whether or not the $(r_2 - r_1)$ hurdle has been surmounted. Assuming this partition can be performed, suppose a loan or gift is arranged from the "have" to the "have not" compartment. If the transfer leaves both compartments between r_3 and r_2 , this will pro-

2. This possibility Solow pointed out in his Figure IX.

vide a relatively rapid way of improving the economy's general welfare. In this connection, also notice that it may pay to insulate market from nonmarket sectors of an economy and leave the non-market sector to its own devices until the more advanced market sector has passed safely into the $r_3 - r_2$ zone.

In summary then, Solow has shown us how to combine population growth, technological change, and the rate of real investment in a manner that permits evaluation of their separate and combined effects on economic development. It may also be that Solow's model will prove useful in a theory of personality development with or without psycho-therapy.

UNIVERSITY OF MINNESOTA

RECENT PUBLICATIONS

- ABEGGLAN, JAMES G. *The Japanese Factory*. Chicago, Illinois: The Free Press. 1958. pp. 142. \$3.50.
- ANDRUS, J. RUSSELL AND MOHAMMED, AZIZALI F. *The Economy of Pakistan*. Stanford, California: Stanford University Press. 1958. pp. 517. \$8.50
- BAUM, WARREN C. *The French Economy and the State*. Princeton, New Jersey: Princeton University Press. 1958. pp. 391. \$7.50.
- BENSON, GEORGE C. S. AND PAYNE, JOHN M. *National Aid to Higher Education*. Washington, D. C.: American Enterprise Association. 1958. pp. 38. \$1.00.
- BLAUG, MARK. *Ricardian Economics — A Historical Study*. New Haven, Connecticut: Yale University Press. 1958. pp. 269. \$5.00.
- BOURNEUF, ALICE. *Norway — The Planned Revival*. Cambridge, Massachusetts: Harvard University Press. 1958. pp. 233. \$5.00.
- BOWMAN, MARY JEAN (EDITOR). *Expectations, Uncertainty, and Business Behavior*. New York: Social Science Research Council. 1958. pp. 202. \$2.00.
- BREWSTER, KINGMAN JR. *Antitrust and American Business Abroad*. New York: McGraw-Hill Book Company, Inc. 1958. pp. 509. \$12.00.
- BULLINGER, CLARENCE E. *Engineering Economy*. New York: McGraw-Hill Book Company, Inc. 1958. pp. 379. \$7.00.
- CHAMBERLAIN, NEIL W. *Labor*. New York: McGraw-Hill Book Company, Inc. 1958. pp. 625. \$7.00.
- CHILD, FRANK C. *The Theory and Practice of Exchange Control in Germany*. The Hague, Netherlands: Martinus Nijhoff. 1958. pp. 241.
- COLM, GERHARD; GEIGER, THEODORE. *The Economy of the American People — Progress, Problems, Prospects*. Washington, D. C.: National Planning Association. 1958. pp. 168. \$2.00.
- COLODNY, ROBERT G. *The Struggle for Madrid: The Central Epic of the Spanish Conflict (1936-37)*. New York: Paine-Whitman Publishers. 1958. pp. 256. \$6.00.
- COMMITTEE FOR ECONOMIC DEVELOPMENT. *Defense Against Inflation*. New York. 1958. pp. 96. \$1.00.
- . *The Problem of National Security*. New York. 1958. pp. 62+. 50 cents.
- CONFERENCE ON RESEARCH IN INCOME AND WEALTH. *A Critique of the United States Income and Product Accounts*. Princeton, New Jersey: Princeton University Press. 1958. pp. 589. \$11.50.
- DAVIDSON, RALPH K.; SMITH, VERNON L.; AND WILEY, JAY W. *Economics: An Analytical Approach*. Homewood, Illinois: Richard D. Irwin, Inc. 1958. pp. 393. \$7.00.
- DAVIES, R. W. *The Development of the Soviet Budgetary System*. New York: Cambridge University Press. 1958. pp. 372. \$8.50.
- DAYSH, G. H. J. (EDITOR). *A Survey of Whitby and the Surrounding Area*. Eton, Windsor, England: The Shakespeare Head Press. 1958. pp. 254.
- DEANE, PHYLLIS (EDITOR). *Bibliography on Income & Wealth*. London: Bowes & Bowes. 1958. pp. 139.
- DEPARTMENT OF ECONOMICS, UNIVERSITY OF LIVERPOOL, G. L. S. SHACKLE (EDITOR). *A New Prospect of Economics*. Liverpool, England: Liverpool University Press. 1958. pp. 498. 30s.
- DUNNING, JOHN H. *American Investment in British Manufacturing Industry*. Fair Lawn, New Jersey: Essential Books. 1958. pp. 365. \$6.30.
- DURAND, JACQUES. *Le Cinema et Son Public*. Paris: Librairie Sirey. 1958. pp. 234.

- ERRERA, MM. J.; SYMON, E.; VAN DER MEULEN, J.; AND VERNAEVE, L. *Euratom — Analyse et Commentaires du Traité*. Bruxelles, Belgium: Librairie Encyclopédique. 1958. pp. 436.
- FARNSWORTH, HELEN C. *Multiple Pricing of American Wheat*. Stanford, California: Food Research Institute, Stanford University. 1958. pp. 50.
- FEYLER, SHERMAN F. *Income Growth with Security*. New York: The Macmillan Company. 1958. pp. 189. \$4.95.
- FIRESTONE, O. J. *Canada's Economic Development 1867-1953*. London: Bowes & Bowes. 1958. pp. 384. 45s.
- FOX, ALAN. *A History of the National Union of Boot and Shoe Operatives 1874-1957*. Oxford, England: Basil Blackwell. 1958. pp. 684.
- GARDLUND, TORSTEN. *The Life of Knut Wicksell*. Stockholm: Almqvist & Wiksell. 1958. pp. 355.
- GRASSINI, F. A. *La Nazionalizzazione dell'Industria del Carbone in Gran Bretagna e in Francia*. Rome: Editrice Studium. 1957. pp. 246.
- HAGEN, EVERETT E. *Handbook for Industry Studies*. Chicago, Illinois: The Free Press. 1958. pp. 89. \$3.50.
- HAGUE, DOUGLAS (EDITOR). *Stability and Progress in the World Economy*. New York: St. Martin's Press. 1958. pp. 267. \$5.00.
- HARRIS, SEYMOUR E. AND OTHERS OF NEW ENGLAND GOVERNORS' TEXTILE COMMITTEE. *New England Textiles and the New England Economy*. Boston, Massachusetts: New England Textile Committee. 1958. \$3.00.
- HARTMANN, GEORGES. *Conjonctures Économiques*. Geneva, Switzerland: Éditions Générales S.A. 1958. pp. 171.
- HENDERSON, HUBERT. *Supply and Demand*. Chicago, Illinois: University of Chicago Press. 1958. pp. 142. \$2.25.
- HENDERSON, JAMES M. *The Efficiency of the Coal Industry*. Cambridge, Massachusetts: Harvard University Press. 1958. pp. 146. \$4.50.
- HOLLAND, DANIEL M. *The Income-Tax Burden on Stockholders*. Princeton, New Jersey: Princeton University Press. 1958. pp. 241. \$5.00.
- HOMAN, PAUL T.; HART, ALBERT GAILORD; SAMETZ, ARNOLD W. *The Economic Order*. New York: Harcourt, Brace and Company. 1958. pp. 839. \$6.95.
- HURTIG, SERGE. *The European Common Market*. New York: Columbia University Press. 1958. 25 cents.
- JONES, F. C. *Hokkaido, Its Present State of Development and Future Prospects*. New York: Oxford University Press. 1958. pp. 146. \$3.40.
- KAPLAN, A. D. H.; DIRLAM, JOEL B.; LANZILLOTTI, ROBERT F. *Pricing in Big Business — A Case Approach*. Washington, D. C.: The Brookings Institution. 1958. pp. 344. \$5.00.
- KENNEDY, WILLIAM F. *Humanist versus Economist*. Berkeley, California. University of California Press. 1958. pp. 96. \$2.00.
- KERR, CLARK; DUNLOP, JOHN T.; REUTHER, WALTER; BUGAS, JOHN S.; COLE, DAVID L.; WITTE, EDWIN E. *U. S. Industrial Relations: The Next Twenty Years*. East Lansing, Michigan: The Michigan State University Press. 1958. pp. 215. \$5.00.
- KINDLEBERGER, CHARLES P. *International Economics, Revised Edition*. Homewood, Illinois: Richard D. Irwin, Inc. 1958. pp. 636. \$7.00.
- KLAMAN, SAUL B. *The Volume of Mortgage Debt in the Postwar Decade*. New York: National Bureau of Economic Research. 1958. pp. 143. \$2.00.
- KNESCHAUREK, FRANCESCO. *Die nationale Buchhaltung*. Zurich: Polygraphischer Verlag. 1958. pp. 212.
- KRUTILLA, JOHN V. AND ECKSTEIN, OTTO. *Multiple Purpose River Development*. Baltimore, Maryland: The Johns Hopkins Press. 1958. pp. 301. \$4.50.
- LABOR AND INDUSTRIAL RELATIONS CENTER, MICHIGAN STATE UNIVERSITY. *The Emerging Environment of Industrial Relations*. East Lansing, Michigan. 1957. pp. 112. \$1.50.

- LANDSBERGER, HENRY A. *Hawthorne Revisited*. Ithaca, New York: Cornell University. 1958. pp. 119. \$3.50.
- LEITER, ROBERT D. *Labor Problems & Industrial Relations*. New York: Barnes & Noble, Inc. 1958. pp. 320. \$1.95.
- LESTER, RICHARD A. *As Unions Mature — An Analysis of the Evolution of American Unionism*. Princeton, New Jersey: Princeton University Press. 1958. pp. 171. \$3.75.
- LEVIN, HARVEY J. (EDITOR). *Business Organization and Public Policy*. New York: Rinehart & Company, Inc. 1958. pp. 550. \$7.00.
- LIBRARY OF CONGRESS. *Marketing Maps of the United States*. Washington, D. C. 1958. pp. 147. \$1.15.
- LUTHER, ERNEST W. *Ethiopia Today*. Stanford, California: Stanford University Press. 1958. pp. 158. \$4.00.
- McKENNA, JOSEPH P. *Intermediate Economic Theory*. New York: Henry Holt and Company (Dryden Press). 1958. pp. 319. \$4.90.
- MEADE, J. E. *The Control of Inflation*. New York: Cambridge University Press. 1958. pp. 52. \$1.00.
- MIERNYK, WILLIAM H.; AND HOROWITZ, MORRIS A. *Engineering Enrollment and Faculty Requirements, 1957-1967*. Boston, Massachusetts: Bureau of Business and Economic Research, Northeastern University. 1958. pp. 59. 25 cents.
- MORLEY, FELIX (EDITOR). *Essays on Individuality*. Philadelphia, Pennsylvania: University of Pennsylvania Press. 1958. pp. 270. \$5.00.
- MUSGRAVE, RICHARD A. AND PEACOCK, ALAN T. (EDITORS). *Classics in the Theory of Public Finance*. New York: The Macmillan Company. 1958. pp. 244. \$6.00.
- MYERS, CHARLES A. *Labor Problems in the Industrialization of India*. Cambridge, Massachusetts: Harvard University Press. 1958. pp. 297. \$6.50.
- MYRDAL, GUNNAR. *Rich Lands and Poor Lands*. New York: Harper & Brothers. 1958. pp. 168. \$2.25.
- NATIONAL BUREAU OF ECONOMIC RESEARCH. *An Appraisal of the 1950 Census Income Data*. Princeton, New Jersey: Princeton University Press. 1958. pp. 450. \$10.00.
- NATIONAL COUNCIL OF INDEPENDENT SCHOOLS. *Preparation of Teachers for Secondary Schools*. Boston, Massachusetts. 1958. pp. 52.
- NETSCHERT, BRUCE C. *The Future Supply of Oil and Gas*. Baltimore, Maryland: The Johns Hopkins Press. 1958. pp. 134. \$3.00.
- NEVIN, EDWARD. *Textbook of Economic Analysis*. New York: St. Martin's Press, Inc. 1958. pp. 422. \$4.50.
- OKUN, BERNARD. *Trends in Birth Rates in the United States Since 1870*. Baltimore, Maryland: The Johns Hopkins Press. 1958. pp. 203+. \$3.50.
- OORT, C. J. *Decreasing Costs as a Problem of Welfare Economics*. Amsterdam, Holland: Drukkerij Holland N. V. 1958. pp. 187. \$1.50.
- PANDIT, V. L. *The Evolution of India*. New York: Oxford University Press. 1958. pp. 46. \$1.00.
- PAPANDREOU, ANDREAS G. *A Test of a Stochastic Theory of Choice*. Berkeley, California: University of California Press. 1957. pp. 18.
- PHELPS, HAROLD A. AND HENDERSON, DAVID. *Population in Its Human Aspects*. New York: Appleton-Century-Crofts, Inc. 1958. pp. 512. \$6.00.
- PINGREE, FRANK. *A Neglected Source of Railroad Profit: Packaged Traffic and How to Transport It Effectively*. Chicago, Illinois: Wolf Management Engineering Company. 1958. pp. 94.
- PIQUET, HOWARD S. *The Trade Agreements Act and the National Interest*. Washington, D. C.: The Brookings Institution. 1958. pp. 61. \$1.25.
- PRITCHARD, LELAND J. *Money and Banking*. Boston, Massachusetts: Houghton Mifflin Company. 1958. pp. 783. \$6.95.

- RADDOCK, MAXWELL C. *Portrait of an American Labor Leader: William L. Hutcheson*. New York: American Institute of Social Science, Inc. 1955. pp. 430. \$5.00.
- RENNE, ROLAND R. *Land Economics*. New York: Harper & Brothers. 1958. pp. 599. \$6.00.
- ROBERTS, B. C. *National Wages Policy in War and Peace*. New York: The Macmillan Company. 1958. pp. 180. \$3.50.
- SAMUELSON, PAUL A. *Economics — An Introductory Analysis*. New York: McGraw-Hill Book Company, Inc. 1958. pp. 810. \$6.75.
- SAPIR, MICHAEL. *The New Role of the Soviets in the World Economy*. New York: Committee for Economic Development. 1958. pp. 64. 50 cents.
- SARGENT, J. R. *British Transport Policy*. New York: Oxford University Press. 1958. pp. 164. \$2.40.
- SHONFIELD, ANDREW. *British Economic Policy Since the War*. Baltimore, Maryland: Penquin Books, Inc. 1958. pp. 288. 85 cents.
- SNIDER, DELBERT A. *Introduction to International Economics, Revised Edition*. Homewood, Illinois: Richard D. Irwin, Inc. 1958. pp. 598. \$7.00.
- STAPLETON, G. BRIAN. *The Wealth of Nigeria*. New York: Oxford University Press. 1958. pp. 228. \$2.60.
- STEINER, W. H.; SHAPIRO, ELI; SOLOMON, EZRA. *Money and Banking, Fourth Edition*. New York: Henry Holt and Company. 1958. pp. 740. \$7.00.
- STEWART, J. D. *British Pressure Groups*. New York: Oxford University Press. 1958. pp. 273. \$4.80.
- TANG, ANTHONY M. *Economic Development in the Southern Piedmont 1860–1950*. Chapel Hill, North Carolina: University of North Carolina Press. 1958. pp. 256. \$6.00.
- TANNENBAUM, ARNOLD S. AND KAHN, ROBERT L. *Participation in Union Locals*. Evanston, Illinois: Row, Peterson and Company. 1958. pp. 275.
- TEITELBAUM, PERRY D. *Nuclear Energy and the U. S. Fuel Economy, 1955–1980*. Washington, D. C.: National Planning Association. 1958. pp. 188. \$3.00.
- THOMAS, BRINLEY (EDITOR). *Economics of International Migration*. New York: St. Martin's Press. 1958. pp. 502. \$9.00.
- THOMAS, DAVID A. *Accelerated Amortization*. Ann Arbor, Michigan: Bureau of Business Research, University of Michigan. 1958. pp. 104. \$5.00.
- TIANO, ANDRÉ; ROCARD, MICHEL; LESIRE-OGREL, HUBERT. *Expériences Françaises d'Action Syndicale Ouvrière*. Paris: Les Éditions Ouvrières. 1956. pp. 435.
- TRAGER, FRANK N. *Building a Welfare State in Burma 1948–1956*. New York: Institute of Pacific Relations. 1958. pp. 118. \$4.00.
- TUCKER, GILBERT M. *Common-Sense Economics*. Harrisburg, Pennsylvania: The Stackpole Company. 1957. pp. 289. \$4.95.
- UNION OF SOUTH AFRICA. *State of the Union 1958*. Cape Town, South Africa: Culemborg Publishers. 1958. pp. 452.
- UNITED NATIONS. *Economic Development of Colombia*. New York: Columbia University Press. 1958. pp. 454. \$4.50.
- . *Economic Survey of Latin America 1956*. New York: Columbia University Press. 1958. pp. 183. \$2.50.
- . *In the Service of Europe*. New York: Columbia University Press. 1958. pp. 48. 35 cents.
- UNITED STATES DEPARTMENT OF AGRICULTURE. *Distributed Lags and Demand Analysis for Agricultural and other Commodities*. Washington, D. C. 1958. pp. 121. 60 cents.
- VANCE, LAWRENCE L. *Theory and Technique of Cost Accounting*. New York: Henry Holt and Company. 1958. pp. 536. \$7.00.
- WALKER, MABEL. *Business Enterprise and The City*. Princeton, New Jersey: Tax Institute, Inc. 1957. pp. 144. \$3.00.

- WESTON, J. FRED. Readings in Finance from *Fortune*. New York: Henry Holt and Company. 1958. pp. 122.
- WOLMAN, WILLIAM. The Development of Manufacturing Industry in the State of Washington. Pullman, Washington: Bureau of Economic and Business Research, State College of Washington. 1958. pp. 208. \$4.00.
- WOODRUFF, W. AND MCGREGOR, L. The Suez Canal and the Australian Economy. New York: Cambridge University Press. 1957. pp. 20. 75 cents.

K Y K L O S

INTERNATIONAL REVIEW FOR SOCIAL SCIENCES

VOL. XI — 1958 — FASC. 4

Public Bureaucracy, Resource Structure, and Economic Development: A Note
Wages and Aggregate Employment
The Means of Competition at Various Stages of Production and Distribution
The Elimination of War — Princeton's New Approach
Economic Science
A Survey of Recent Contributions to the Economics of Development
Common Frontiers of the Social Sciences

Joseph J. Spengler
Franz Gehrels
Gösta Mickwitz
C. A. W. Manning
Robert Dorfman
A. Bonné
Thomas J. Pressly

BOOK REVIEWS — TITLES OF NEW BOOKS

All articles are followed by a summary in English, French and German. Published quarterly. Annual subscription Sw.Fr. 26. — (U.S. \$6.00), postage included. Order from your bookstore or the publisher: Kyklos-Verlag, Postfach 20 365, Basel 2, Switzerland.

THE ECONOMIC JOURNAL

The Quarterly Journal of the Royal Economic Society

No. 271

September 1958

Vol. LXVIII

I. Articles

The Place of Economics and Allied Subjects in the Curriculum
Inflation and the British Economy
Welfare Propositions in Terms of Consistency and Expanded Choice
Professor Hansen on America's Economic Revolution
Import Restriction as an Instrument of Balance-of-Payments Policy

Sir Alexander Carr-Saunders
A. J. Brown
K. Lancaster
K. K. Kurihara

Capital Theory and Economic Growth
Differential Profits Tax

M. F. W. Hemming and W. M. Corden
K. Ara
T. Balogh

II. Reviews

III. Notes and Memoranda

IV. Recent Periodicals and New Books

Macmillan and Co., Ltd., London, W.C. 2

Application for Fellowship to

Royal Economic Society, 21 Bentinck Street, London, W. 1

Annual Subscription £1 10s.

Single Copy Ten Shillings

PUBLIC FINANCE / FINANCES PUBLIQUES

International quarterly journal devoted to the study of fiscal policy and related problems
Publisher and Editor in Chief: DR. J. A. MONOD DE FROIDEVILLE
Emmalaan 2, Haarlem, Netherlands

VOLUME XIII

1958

NUMBER 2

- Ugo Papi, Rome.....European Economic Cooperation
Alan T. Peacock and I. G. Stewart,
Edinburgh, Great Britain.....Fiscal Policy and the Composition of Govern-
ment Purchases
N. A. Cox-George, Fourah Bay College,
Sierra Leone.....Studies in Finance and Development — the Gold
Cost (Ghana) Experience 1914-1918

Book Reviews

Contributions in French or German are followed by an English Summary
Subscription Price for 1958: \$10.50, postage free.

REVIEW OF ECONOMIC STUDIES

Vol. XXV (3)

JUNE 1958

No. 68

CONTENTS

Tariffs, Retaliation and the Elasticity of Demand for Imports, *W. M. Gorman*; The Wicksell Effect, *Edward Osborn*; Utility and the "Ordinalist Fallacy," *W. F. Armstrong*; The Predictive Accuracy of the Marginal Productivity Theory of Wages, *J. L. Stein*; The Relation between the Accelerator and the Capital Output Ratio, *Alfred E. Ott*; Demand Theory without a Utility Index — A Comment, *P. K. Newman* and *R. C. Read*; Further Comments, *Lionel McKenzie*; Immiserising Growth: A Geometrical Note, *Jagdish Bhagwati*; Risk Bearing and Income Taxation, *Nicholas Kaldor*.

Annual Subscription, 21s. per Vol. (3 issues), post free. U. S. A. and Canada, \$4.00 per Vol. Back Number Rates on Application to Secretary.

With Vol. XVI (1), the Review reverted to three issues per volume. The new subscription rates appear above. Editorial communications should be sent to the Editor, Review of Economic Studies, Nuffield College, Oxford. Articles from the U. S. A. and Canada should be sent to P. A. Samuelson, Massachusetts Institute of Technology, Cambridge, Mass. All other communications to the Secretary, Review of Economic Studies, 4 Trumpington Street, Cambridge, England. Cheques should be made payable to the Review of Economic Studies.

REVUE DE SCIENCE FINANCIÈRE

THE FRENCH REVIEW OF PUBLIC FINANCE

Number IV. October–December 1958

The Greater Paris: Economic, Administrative and Financial Problems *P. Sudreau (Paris)*
Taxation of Real Property: Reexamination of Its Functions and Forms *N. D. Ganjei (New York)*
Financial Rules of International Organizations: the Principle of a Single Budget

The Foreign Currency Standard: Scope and Limitations *C.-A. Colliard (Grenoble)*
Public Investments in the French Economy, 1947–1958 *P. Wolf (Paris)*

Survey of Foreign Finance Survey of French Public Credit *P. Martinot (Toulouse)*
Book Reviews Bibliography French and International Financial Documents

Notes and Memoranda:

Regional Economic Development in Poland: Some of Its Aspects *Th. Lutoborski (Warsaw)*
The Bank of International Settlements: Financial Highlights of the 28th Report

H. Laufenburger (Paris)

The Review is published quarterly. Yearly subscription abroad 4,000 French francs or 9.50 U.S.A. dollars. Orders can be placed to Pichon and Durand-Auzias, 20 rue Soufflot, Paris (V^e), and through Stechert-Hafner, 31 East 10th Street, New York, N. Y.

THE JOURNAL OF POLITICAL ECONOMY

Edited by ALBERT REES in Cooperation with the Other Members of the
Department of Economics of the University of Chicago

The October 1958 issue will contain:

Monetary Reform and Monetary Equilibrium, *Boris P. Peseck*; Economies of Scale in Financial Enterprises, *Roy J. Hensley*; A Theory of Foreign-Exchange Speculation under a Floating Exchange System, *S. C. Tsiang*; Research Costs and Social Returns: Hybrid Corn and Related Innovations, *Zvi Griliches*; An Economic Analysis of Contributions under the Income Tax Laws, *Floyd E. Gillis* and *Vernon L. Smith*; The Economics of Slavery in the Ante Bellum South: A Comment, *Douglas F. Dowd*; Reply, *Alfred H. Conrad* and *John R. Meyer*; The Corporate Income Tax in the Long Run: A Comment, *Arnold Zellner*; Reply, *M. A. Adelman*; Rejoinder, *Arnold Zellner*. Book Reviews and Books Received

FOUNDED IN 1892. Published bimonthly: February, April, June, August, October, December. Subscription rate \$6.00 per volume. Single copy \$1.50. Special rates: \$5.00 per volume for faculty members teaching business or economics, \$3.00 per volume for graduate students in business or economics. Canadian postage 40 cents. Foreign postage 75 cents.

THE UNIVERSITY OF CHICAGO PRESS, 5750 Ellis Avenue, Chicago 37, Ill.

METROECONOMICA

Volume X

INTERNATIONAL REVIEW OF ECONOMICS

Number 3

Macrodynamic policy-making

Histoire de l'économie mathématique avant Cournot

Business investment and marginal cost (2d part)

A restatement of the Conditions for identifiability in

complete systems of linear difference equations

The aggregate supply function

The problem of intrinsic complementarity and separability of goods

The valuation of random income streams

M. J. Bogaard and H. Theil

G. H. Bousquet

H. J. Cunningham

H. S. Konijn

A. L. Marty

M. Morishima

A. Roy

Board of Editors: C. Bresciani Turrone, L. Dupriez, E. Fossati (*Managing Editor*), R. Frisch, W. Leontief, G. Lutfalla, A. Mahr, J. Marschak, G. U. Papi, R. Roy, E. Schneider, G. L. S. Shackle, R. Stone, J. Tinbergen, G. Tintner, F. Zeuthen.

Published every four months. Subscription rate \$7.00 per year (two years \$12.00). Communications to the editor to be addressed to *Professor Eraldo Fossati*, director of the Laboratorio di economia Vilfredo Pareto, Università, Genova, (Italy). Communications regarding subscriptions should be sent to *Licínio Cappelli*, Corso 12, Trieste or *G. E. Stechert-Hafner Inc.*, 31 East Tenth Street, New York 3, N. Y., U.S.A.

Good earnings for the telephone company have a way of being good for the customer, too



Everybody has problems. One of ours is the rather widespread belief that the sure way to low telephone rates is to keep the company's earnings as low as possible.

Attractive as this may seem at the moment, it has distinct hazards for the telephone user. Handcuffing the company limits progress and long-range economies, and will lead to poorer service at a higher price than the customer would otherwise have to pay.

This fact is receiving increasing recognition by the commissions which regulate telephone rates and earnings. But it is not something that concerns the commissions alone.

It needs your understanding, too.

Regulation can only work best in the interest of all when it is free to act in the interest of all on the basis of economic facts.

Authorizing good earnings for the telephone company requires wisdom and foresight and sometimes calls for a high degree of political courage. Such action, however, in the long run will return the greatest value to the public.

Fortunately, from the standpoint of the individual telephone user, our increased rate needs, where required, are small. They usually come to not more than a few cents a day on the average customer's bill.

BELL TELEPHONE SYSTEM



American Economic Review

CONTENTS

VOLUME XLVIII

SEPTEMBER 1958

NUMBER 4

ARTICLES

Soviet and American Inventory-Output Ration

R. W. Campbell

The Firm in Illyria: Market Syndicalism

Benjamin Ward

Progressive Taxation in an Inflationary Economy

E. J. Misham and L. A. Dicks-Mireaux

The Size Distribution of Business Firms

H. A. Simon and C. P. Bonini

The Constancy of Relative Shares

R. M. Solow

Economic Problems in Air Force Logistics

Horst Mendershausen

COMMUNICATIONS

Consumer Loans and General Credit Conditions

Paul Smith

The Income Tax and Deductible Costs

J. N. Morgan

Korea's Monetary and Fiscal Systems:

Comment

Louis de Alessi and Robert Evans

Comment

Sei-Young Park

Reply

C. D. Campbell and Gordon Tullock

The Tax Element in Soviet Bonds:

Comment

Morris Bornstein

Reply

F. D. Holzman

The AMERICAN ECONOMIC REVIEW, a quarterly, is the official publication of the American Economic Association and is sent to all members. The annual dues are \$6.00. Address editorial communications to Dr. Bernard F. Haley, Editor, American Economic Review, Stanford University, Room 220, Stanford, California; for information concerning other publications and activities of the Association, communicate with the Secretary-Treasurer, Dr. James Washington Bell, American Economic Association, Northwestern University, Evanston, Illinois. Send for information booklet.

POPULATION GROWTH AND ECONOMIC DEVELOPMENT IN LOW-INCOME COUNTRIES

By ANSLEY J. COALE and EDGAR M. HOOVER

Here is an important study of the alternatives confronting low-income countries with respect to population growth that shows the nature and extent of economic gains resulting from fertility reduction. India is used as a case study in this book that has direct implications for all low-income agrarian areas entering into a program of economic development.

400 pages. Tables & charts. \$8.50

Order from your bookstore, or

PRINCETON UNIVERSITY PRESS, Princeton, New Jersey

Economic Theory and Western European Integration

TIBOR SCITOVSKY. Some of the general principles of economic theory are applied to the particular problems of the economic union of Western Europe. \$3.50

The Staple Food Economies of Western Tropical Africa

BRUCE F. JOHNSTON. Considers all aspects of the group of starchy staple foods that are of predominant importance in African diets. Maps and charts. *A publication of the Food Research Institute.* \$6.00

Studies in Linear and Non-Linear Programming

KENNETH J. ARROW, LEONID HURWICZ, and HIROFUMI UZAWA. The problems dealt with have their origins in practical questions of business and economic policy. *Stanford Mathematical Studies in the Social Sciences, II.* \$7.50

The Uniting of Europe

ERNST B. HAAS. Political, Social, and Economic Forces, 1950–1957. Describes the aims and operations of the European Coal and Steel Community during the first five years of its existence. \$8.00


French West Africa

VIRGINIA THOMPSON and RICHARD ADLOFF. In the first comprehensive analysis of this area in almost 20 years, the important changes in the economic structure are emphasized. Photographs and maps. \$8.50

Political Change in Latin America

JOHN J. JOHNSON. A study of the continuing economic revolution in Latin America that has brought to the surface the politically ambitious urban middle groups. \$5.00

***Stanford University Press
Stanford, California***



A Check List of Books on Economics Published During 1958

Bassie — ECONOMIC FORECASTING
702 pages, \$8.75

Chamberlain — LABOR
625 pages, \$7.00

Chamberlain — SOURCEBOOK ON LABOR
1,104 pages, \$9.75

Dusenberry — BUSINESS CYCLES AND ECONOMIC GROWTH
341 pages, \$6.50

Henderson and Quandt — MICROECONOMIC THEORY
291 pages, \$7.50

Kindleberger — ECONOMIC DEVELOPMENT
325 pages, \$6.50

Samuelson — ECONOMICS, 4th edition
832 pages, \$6.75

Samuelson-Bishop-Coleman — READINGS IN ECONOMICS
473 pages, \$4.50; paper ed., \$3.25

. . . and coming in early 1959

Allen-Buchanan-Colberg — PRICES, INCOME AND PUBLIC POLICY
2/e (Jan.)

Clough — THE ECONOMIC DEVELOPMENT OF WESTERN
CIVILIZATION (Jan.)

Coppock — ECONOMICS OF THE BUSINESS FIRM (March)

Harbison-Ibrahim — HUMAN RESOURCES FOR EGYPTIAN
ENTERPRISE (Jan.)

Kemmerer-Jones — AMERICAN ECONOMIC HISTORY (May)

Lewis — ANALYZING BUSINESS CONDITIONS (March)

Musgrave — THE THEORY OF PUBLIC FINANCE (Jan.)

Schlaifer — PROBABILITY AND STATISTICS FOR BUSINESS
DECISIONS (March)

Valavanis — THE ABC OF ECONOMETRICS (May)

Send for copies on approval

McGraw-Hill Book Company, Inc.

330 West 42nd Street

New York 36, N.Y.

